

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 16:34:22 ; Search time 69.24 Seconds  
(without alignments)  
1084.341 Million cell updates/sec

Title: US-08-816-011f-63  
Perfect score: 2294  
Sequence: 1 MVIINRNTYAVEQEAFFRD.....LTGGLWPVVIEHFVKPSIL 434

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 segs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPREMBL\_19:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mmc:\*
- 8: sp\_organelle:\*
- 9: sp\_phase:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2265	98.7	434	5 Q19495	Q19495 caenorhabdi
2	802	35.0	418	5 Q9U3L8	Q9U3L8 caenorhabdi
3	716.5	31.2	461	5 Q23599	Q23599 caenorhabdi
4	706.5	30.8	450	5 Q93871	Q93871 caenorhabdi
5	645	28.1	458	5 Q9VMB6	Q9VMB6 drosophila
6	574.5	25.0	800	5 Q21145	Q21145 caenorhabdi
7	367.5	16.0	475	11 Q99P65	Q99P65 mus musculu
8	343.5	15.0	456	4 Q9UJY2	Q9UJY2 homo sapien
9	336.5	14.7	456	6 Q9BG85	Q9BG85 oryctolagus
10	328.5	14.3	456	11 Q9JIT8	Q9JIT8 mus musculu
11	327.5	14.3	475	4 Q9BZD2	Q9BZD2 homo sapien
12	320.5	14.0	428	10 Q9M5X9	Q9M5X9 arabisopsis
13	313.5	13.7	428	10 Q64603	Q64603 arabisopsis
14	302.5	13.2	458	11 Q9DBT8	Q9DBT8 mus musculu
15	302.5	13.2	458	11 Q9JHFO	Q9JHFO mus musculu
16	301.5	13.1	460	11 Q9JIMI	Q9JIMI mus musculu

17	301.5	13.1	460	11 Q99K84	Q99K84 mus musculu
18	291.5	12.7	476	5 Q9NH21	Q9NH21 drosophila
19	291.5	12.7	476	5 Q95RN3	Q95RN3 drosophila
20	261.5	11.4	397	4 Q9NUS9	Q9NUS9 homo sapien
21	249.5	10.9	408	10 Q9FWI1	Q9FWI1 arabisopsis
22	221.5	9.7	586	5 Q9VU20	Q9VU20 drosophila
23	217.5	9.5	404	5 Q9VPP0	Q9VPP0 drosophila
24	197	8.6	497	5 Q9GTP5	Q9GTP5 crithidia f
25	195.5	8.5	415	6 Q9BG84	Q9BG84 oryctolagus
26	195	8.5	384	5 Q93690	Q93690 caenorhabdi
27	195	8.5	491	5 Q76343	Q76343 leishmania
28	187	8.2	491	5 Q76269	Q76269 leishmania
29	185.5	8.1	285	4 Q9BWI2	Q9BWI2 homo sapien
30	182.5	8.0	668	5 Q961M5	Q961M5 drosophila
31	182	7.9	361	4 Q96PB2	Q96PB2 homo sapien
32	181	7.9	301	4 Q96R00	Q96R00 homo sapien
33	181	7.9	501	5 Q9N9R1	Q9N9R1 leishmania
34	181	7.9	549	5 Q95YT5	Q95YT5 leishmania
35	180	7.8	143	5 Q20396	Q20396 caenorhabdi
36	170	7.4	499	5 Q9NBV4	Q9NBV4 leishmania
37	161	7.0	417	10 Q944P0	Q944P0 arabisopsis
38	151	6.6	463	5 Q9U763	Q9U763 trypanosoma
39	147	6.4	417	10 Q9SR64	Q9SR64 arabisopsis
40	144	6.3	418	10 Q9M0Y3	Q9M0Y3 arabisopsis
41	142	6.2	462	5 Q9TVQ1	Q9TVQ1 toxoplasma
42	142	6.2	463	5 Q95Z04	Q95Z04 trypanosoma
43	141	6.1	417	10 Q944N9	Q944N9 arabisopsis
44	138.5	6.0	418	10 Q9M0Y2	Q9M0Y2 arabisopsis
45	136	5.9	463	5 Q9Y010	Q9Y010 trypanosoma

ALIGNMENTS

RESULT 1  
Q19495 PRELIMINARY; PRT; 434 AA.  
AC Q19495;  
DT 01-NOV-1996 (TREMURel. 01, Created)  
DT 01-NOV-1996 (TREMURel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMURel. 19, Last annotation update)  
DE HYPOTHETICAL 49.3 KDA PROTEIN.  
GN F16H11.3.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
RT investigating biology. The C. elegans Sequencing Consortium.;"  
RL Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Wu X.;  
RT "The sequence of C. elegans cosmid F16H11.1;"  
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BRISTOL N2;  
RA Waterston R.;  
RT "Direct Submission.;"  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U55376; AAA98003.1; -  
DR InterPro: IPR002259; DER\_Nucleoside\_tran.  
DR Pfam: PF01733; Nucleoside\_tran; 1.  
DR ProDom: PD005103; DER\_Nucleoside\_tran; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 434 AA; 49317 MW; 669478D51833965 CRC64;

X not good date

Query Match 98.7%; Score 2265; DB 5; Length 434;  
Best Local Similarity 99.1%; Pred. No. 2e-177; Mismatches 4; Indels 0; Gaps 0;  
Matches 430; Conservative 0;

QY 1 MVIINRNTYAVEQEAFFPRDKYNIYVWLVILVGFVLLPWNMFITIAPEYYVNVYVFKPDG 60  
DB 1 MVIINRNTYAVEQEAFFPRDKYNIYVWLVILVGFVLLPWNMFITIAPEYYVNVYVFKPDG 60

QY 61 VETWYSKEFMGSLTIGSQLPNASINVFNLFLIIAGPLIYRVFAPVCFNIVNLTIILIVI 120  
DB 61 VETWYSKEFMGSLTIGSQLPNASINVFNLFLIIAGPLIYRVFAPVCFNIVNLTIILIVI 120

QY 121 VLEPTEDSMWFFWVTLGMAISINFGYLSVLLVCAIALEFFITKQDFYHHQKGMIREKA 180  
DB 121 VLEPTEDSMWFFWVTLGMAISINFGYLSVLLVCAIALEFFITKQDFYHHQKGMIREKA 180

QY 181 TVVKGIVTYFLNDEPKLVAIVYFGISLVLLVCAIALEFFITKQDFYHHQKGMIREKA 240  
DB 181 TVVKGIVTYFLNDEPKLVAIVYFGISLVLLVCAIALEFFITKQDFYHHQKGMIREKA 240

QY 241 ETRDPSILMTTFTNCYQGLFNWFCFAVLTITFPVMMVTTRGDSGFLNKMSENDEI 300  
DB 241 ETRDPSILMTTFTNCYQGLFNWFCFAVLTITFPVMMVTTRGDSGFLNKMSENDEI 300

QY 301 YTLTSLFVNLFAAIGSIVASKIHWPTRYLKFAIILRALFIPFFFCNRYRVQTRAYPV 360  
DB 301 YTLTSLFVNLFAAIGSIVASKIHWPTRYLKFAIILRALFIPFFFCNRYRVQTRAYPV 360

QY 361 PFESTDIFVIGGIAMSFHGYLSALAMGYTPNVVPSHSRFAAQLSVCTLMVGLTGLW 420  
DB 361 PFESTDIFVIGGIAMSFHGYLSALAMGYTPNVVPSHSRFAAQLSVCTLMVGLTGLW 420

QY 421 PVVIEHFVDKPSIL 434  
DB 421 AVVIEHFVDKPSIL 434

RESULT 2  
QY 3 Q9U3L8 PRELIMINARY; PRT; 418 AA.  
AC Q9U3L8;  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
DE C47A4.2 PROTEIN.  
GN C47A4.2.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Mortimore B.J.;  
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99069613; PubMed=9851916;  
RA none;  
RT "Genome sequence of the nematode C.elegans: A platform for  
investigating biology.";  
RL Science 282:2012-2018(1998).  
DR EMBL; Z82263; CAB62793.1; -.  
DR InterPro; IPR002259; DER\_Nucleoside\_tran.  
DR Pfam; PF01733; Nucleoside\_tran; 1.  
DR ProDom; PD005103; DER\_Nucleoside\_tran; 2.  
SQ SEQUENCE 418 AA; 46333 MW; 5CBDD27E069FC8D CRC64;

Query Match 35.0%; Score 802; DB 5; Length 418;  
Best Local Similarity 37.6%; Pred. No. 9.5e-58;  
Matches 156; Conservative 84; Mismatches 135; Indels 40; Gaps 2;

QY 3 IINRNTYAVEQEAFFPRDKYNIYVWLVILVGFVLLPWNMFITIAPEYYVNVYVFKPDGVE 62  
DB 30 IVDGADSDDEHNEIPEDKGRIVFWIILLNGICVLLPWNMFITIAPEYYVNVYVFKPDGVE 89

QY 63 TWYSKEFMGSLTIGSQLPNASINVFNLFLIIAGPLIYRVFAPVCFNIVNLTIILIVI 122  
DB 90 THYADSFQSGVAAQVQPNLVAIINVLNIRGLPLIYRVLAFLAFNLLIIVILALVIFQ 149

QY 123 EPTEDSMWFFWVTLGMAISINFGYLSVLLVCAIALEFFITKQDFYHHQKGMIREKAET 182  
DB 150 QESQARMEFYIVSLIIVMAMNASGLYQNSFEFGMAADFPKYSNAVIGTICGTTSTV 209

QY 183 VKIGVTYFLNDEPKLVAIVYFGISLVLLVCAIALEFFITKQDFYHHQKGMIREKAET 242  
DB 210 LAIVATLAFSQAENVALIYFGISLVLLVCAIALEFFITKQDFYHHQKGMIREKAET 250

QY 243 DRPSPSILMTTFTNCYQGLFNWFCFAVLTITFPVMMVTTRGDSGFLNKMSENDEIYT 302  
DB 251 -----CWLOICVFLVYFVSLVFPVLAGFQPGYTFPN-----DVYA 289

QY 303 LTLTSLFVNLFAAIGSIVASKIHWPTRYLKFAIILRALFIPFFFCNRYRVQTRAYPV 362  
DB 290 GIAVFLNFFFAAGVAAATFVTFPGRLLIIVPCVIRLLIFIPFFFMFSNLYPHSRMTGVLF 349

QY 363 ESTDIFVIGGIAMSFHGYLSALAMGYTPNVVPSHSRFAAQLSVCTLMVGLTGL 417  
DB 350 TNEWIFFGNTLLAFTSGTSSGLMMYTPRCVCPPEYSKLAGOVSAISLVIGITAG 404

RESULT 3  
QY 3 Q23599 PRELIMINARY; PRT; 461 AA.  
AC Q23599;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE ZK809.4 PROTEIN.  
GN ZK809.4.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Dobson R.;  
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99069613; PubMed=9851916;  
RA none;  
RT "Genome sequence of the nematode C.elegans: A platform for  
investigating biology.";  
RL Science 282:2012-2018(1998).  
DR EMBL; Z68303; CAA92642.1; -.  
DR InterPro; IPR002259; DER\_Nucleoside\_tran.  
DR Pfam; PF01733; Nucleoside\_tran; 1.  
DR ProDom; PD005103; DER\_Nucleoside\_tran; 1.  
SQ SEQUENCE 461 AA; 52050 MW; CD0F115AF6A12A76 CRC64;

Query Match 31.2%; Score 716.5; DB 5; Length 461;  
Best Local Similarity 35.8%; Pred. No. 1e-50;  
Matches 163; Conservative 86; Mismatches 167; Indels 39; Gaps 9;

QY 7 SNTYAVEQEA-FPRDKYNIYVWLVILVGFVLLPWNMFITIAPEYYVNVYVFKPD----- 59  
DB 11 NKTKKVEEPEPEKGNLVFVILLHGIGTILMPNMLITISDYFESYKMLANLTIIDMD 70

QY 60 -----GVETWYSKEFMGSLTIGSQLPNASINVFNLFLIIAGPLIYRVFAPVCFNIVNLTI 114  
DB 71 TGVWTGYPTVYSSNFQSTIASQVFNLLNLLNIFVIVKGLASRI--TVGLSIVAVCV 128

```
QY 115 ILILVIVLEPTEDSMWFFWVTLTGMATSI NFNGLYENSVYGVGDFFPHYTGALLIGN 174
Db 129 ITTMMFIYVETSTWLTGFTTLITIIIVLNGANGVQNSIFGLASELPFKYTNVAILGN 188
QY 175 ICGLLITVWKI--GVYFLNDEPKLVAIVFGISLVILLVCATALFITKQDFHYHQ 231
Db 189 LCGFTVLLSMTKAVTRNILD R-----SFAYFSIALITLVFCFISHLKKQRFQYYST 244
QY 232 KGMIREK--AETDRSPSILMTTNCYQGLPNVWFCFAVTLTTP-VNMTVTTGRDS 287
Db 245 RAERQRKNDEAVDSEKGVANYIAITFEAFQQLINVLVFFVTLISIPGVMMYVKDEKG 304
QY 288 G-----FLNKIMSENDEI-----YTLTSLFVNLFAAIGSIVASKIHWPTPRYLK 333
Db 305 GVIDFPLPRQFFINSLTHNREFSEKVFMDVTTFLQFNVFAFGISIVAGRKQWPAPNKLW 364
QY 334 FAITLRALFIPFEFCNVRVQTRAYPVFFESTDIFVIGGIMATSHGYSLSALAMGYTPNV 393
Db 365 IPVYLRLLYIPFFIFCNLYPSTRPLPVFFESTWLFVIAAASMSFGSGYFGLAMMYTSKT 424
QY 394 VPSHYSRFAQLSVCTLMVGLLTGLPVPVIEHFV 428
Db 425 VDPKSAQVAGNMAGFELISGIVSLIETWIKMVV 459
RESULT 4
Q93871 ID Q93871 PRELIMINARY; PRT; 450 AA.
AC Q93871
DT 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE K09A9.3 PROTEIN.
GN K09A9.3
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN SEQUENCE FROM N.A.
RA Swinburne J.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z79601; CAB01882.1; -.
DR InterPro; IPR002259; DER_Nucleoside_tran.
DR Pfam; PF01733; Nucleoside_tran; 1.
DR ProDom; PD005103; DER_Nucleoside_tran; 1.
SQ SEQUENCE 450 AA; 50495 MW; CF3973D5EE517909 CRC64;
Query Match 30.8%; Score 706.5; DB 5; Length 450;
Best Local Similarity 36.0%; Pred No. 6.7e-50;
Matches 157; Conservative 88; Mismatches 164; Indels 27; Gaps 8;
QY 12 VEQEA-PPRKYNIYVWLVLGVGLVPLNMFITIAPEYVNVWF-----KPD 59
Db 21 VEEETPEPKGNLVFVILLHGIGTLMPNMLTISVDYFESYKMLANSTIDMDTKVT 80
QY 60 GVEWYSEKFGSLTISQQLPNASINVENFLITAGLPIYVFPVCFNIVNLITILV 119
Db 81 GDPTVYSNFSQFOTIASQVFNLLNLNIFVYKGLAGRI--TVGLSIVAVCVITMI 138
QY 120 IVLEPTEDSMWFFWVTLTGMATSI NFNGLYENSVYGVGDFFPHYTGALLIGNICGLL 179
Db 139 FIYVETSTWLTGFTTLITIIIVLNGANGVQNSIFGLASELPFKYTNVAILGNLCGTF 198
QY 180 ITVVKI---GVYFLNDEPKLVAIVFGISLVILLVCALIALEFFITKQDFHYHQKMEI 236
```

```
Db 199 VTLSSMTKAVTRNILD R-----SFAYFSIALITLVFCFISHLKKQRFQYSTRABEQ 254
QY 237 R---EKAETDRSPSILMTTNCYQGLPNVWFCFAVTLTTP-VNMTVTTGRDGGFLNK 292
Db 255 RAKNEEAADNEGKANYIATFEKAPQQLINVLVFFVTLISIPGVMMYVKDEKKGTFDF 314
QY 293 IMSNDEIYTLTSLFVNLFAAIGSIVASKIHWPTPRYLKFAITLRALFIPFEFCNVR 352
Db 315 PLQN--YFMDVTTFLQFNVFAFGISIVAGRKQWPAPNKLWIPVYLRLLYIPFFIFCNV 372
QY 353 VQTRAYPVFFESTDIFVIGGIMATSHGYSLSALAMGYTPNVVPSHYSRFAQLSVCTLMV 412
Db 373 PETRTLVPFEESTWLFVIAAASMSFGSGYFGLSNMYSKTVDPSKAQVAGNMAGFELIS 432
QY 413 GLLTGGLPVPVIEHFV 428
Db 433 GIVSGLIFTWIKFVV 448
RESULT 5
Q9VMB6 ID Q9VMB6 PRELIMINARY; PRT; 458 AA.
AC Q9VMB6;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE CG11045 PROTEIN (CH12067P).
GN BEST:LD04971 OR CG11045.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Eohydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.M., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrer A., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Hariri N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclib J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinekt K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Zhang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
```

```
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of Drosophila melanogaster.";  
RL  
RN  
RP  
RC STRAIN=Y, CN BW SP;  
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Muncall C.J.,  
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,  
RA Yu C., Lewis S.E., Rubin G.M., Ceiniker S.;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AEO03614; AAF52405.1; -  
DR EMBL; AY058347; AAL13576.1; -  
DR FlyBase; FBgn0026585; BEST:LD04971.  
DR InterPro; IPR002259; DER_Nucleoside_tran.  
DR Pfam; PF01733; Nucleoside_tran; 1.  
DR PRINTS; PRO1130; DERENTRNSPRT.  
DR ProDom; PD005103; DER_Nucleoside_tran; 1.  
SQ SEQUENCE 458 AA; 51324 MW; 72D36D64CE0D4A2E CRC64;  
  
Query Match 28.1%; Score 645; DB 5; Length 458;  
Best Local Similarity 32.9%; Pred. No. 7.4e-45;  
Matches 138; Conservative 83; Mismatches 178; Indels 20; Gaps 7;  
  
Qy 18 PRDKNIVYVLVILVGVGLLPNNMFITIAPEYVYNYWFKPD---GVETWYSKEFMGSLT 74  
Db 51 PRDKFLIVFFILFLHGVGLPNNMFIT-AKSYFEDFKGPNNTVATVSTRTFMQNG 109  
Qy 75 IGSQPNASINVFNLIIAGPLIYRVPVCFNIVNLTIILIVIVLEPTEDSMW--- 131  
Db 110 FASQIPNLVFNWLAIFVNFGLDTRIVYSIIFEMVILLVITLAML-----DSSQWPGV 164  
Qy 132 FFWTLGMATINSNGLYENSVGVGDPPHTYIGALLIGNNCGLLITVVKIGVTYFL 191  
Db 165 FFWTMTWICVLVNCNGIYQNTIYGIVASLPKYTGAVLVGNSGCGTTAMAL-ICGEI 223  
Qy 192 NDEPKLVAIVFGSLVILLCAIALRFTTKQDFVYHHQGMETREKAETDRSPSLTW 251  
Db 224 FSSKRTSAIYFVVAIVLLLCFDYFALPLNKFRHYETISRSEKSKSAQLNVPYW 283  
Qy 252 TTFNCGYQLFNWVFCFAVTLTIFFVMMVTTRGDSGLFKIMSENDEIYTLTSLFVN 311  
Db 284 QIFKKAAPQLENIELTFRVTLVSVEPAIQSNVHRSDPNFV-----VGPDYFLVTCFAFN 338  
Qy 312 LFAAGISIVASKIHWPPRYLKALILRALRPIFFFCNY--RVQTRAYPVFFESTDIFV 369  
Db 339 VFAMLGSLTTSWQWPGPRFLVVPVVLRLAFIPLFVCMCNYPVPPDSVRSVLAFFIENDWYV 398  
Qy 370 IGGTAMSFSGYLSALAMGYTPNVVPSHSYRFAAQLSVCTLMVGLLTGGLWPVIEHFV 428  
Db 399 GIGANAYSSGYLSLGMYPQVHTKYQTAGMYAAMLITGFSGLVSLYLGPFV 457  
  
RESULT 6  
Q21145 PRELIMINARY; PRT; 800 AA.  
AC Q21145;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE K02E11.1 PROTEIN.  
GN K02E11.1  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA McMurray A.A.;  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
RN [2]
```

```
RP SEQUENCE FROM N.A.  
RX MEDLINE=99069613; PubMed=9851916;  
RA none;  
RT "Genome sequence of the nematode C.elegans: A platform for  
RT investigating biology.";  
RL Science 282:2012-2018(1998).  
DR EMBL; Z77665; CAB01223.1; -  
DR InterPro; IPR002259; DER_Nucleoside_tran.  
DR Pfam; PF01733; Nucleoside_tran; 1.  
DR ProDom; PD005103; DER_Nucleoside_tran; 1.  
SQ SEQUENCE 800 AA; 90383 MW; E05D51129E4BF45 CRC64;  
  
Query Match 25.0%; Score 574.5; DB 5; Length 800;  
Best Local Similarity 32.7%; Pred. No. 7.8e-39;  
Matches 134; Conservative 88; Mismatches 165; Indels 23; Gaps 10;  
  
Qy 19 RDKNIVYVLVILVGVGLLPNNMFITIAPEYVYNYWFK-PDGVETWYSKEFMGSLTIGS 77  
Db 388 KDLGNVFFIFMFEFGALLPNNMFNISFDYTMFKLSADGNATWYSSNFQNSMTISA 447  
Qy 78 QLPNASINVFNLIIAGPLIYRVPVCFNIVNLTIILIVIVLEPTEDSMWFWNTL 137  
Db 448 QIPSLVFSVINIFIAVKGDLTRGM--KICLIWVQLMVITVYFIIDTSTWIATESMLTL 505  
Qy 138 GMATINSNGLYENSVGVGDPPHTYIGALLIGNNCGLLITVVKIGVTYFLNDEPKL 197  
Db 506 GTIVVLNANGLFQNSMFGLASPPFFKYTNVAVIIGONFCGTAVVLSM-LTKAASDDVQM 564  
Qy 198 VAIYFGISLVILLCAIALRFTTKQDFVYHHQGMETREKAETDRSPSLTW--- 253  
Db 565 RASLFFGLSSVAVVVCFTLLNFKRLAFY----KFGILRTSSQSDDEGIS-SWESVKLA 619  
Qy 254 FTNCGYQLFNWVFCFAVTLTIFFVMMVTTRGDSGLFKIMSENDEIYTLTSLFVNLF 313  
Db 620 FEKSMQFANIFLVFFVTLALFPNVCMYVKDAKKGELSHFVVP-EKYMVDVVYTLFNLF 678  
Qy 314 AAGISIVASKIHWPPRYLKALILRALRPIFFFCNY--RAYPVFFESTDIFVIG 371  
Db 679 AFLGSLMANWTRFQPNVTWICVAARFWMFYFPAANYHPMDFPRAYPVLESTWLFN 738  
Qy 372 GIAMSFSGYLSALAMGYTPNVVPSH-----YSRFAAQLSVCTLMVGLLTG 417  
Db 739 ICIPALTSGYLSLIMYAPR---SHEDPKIQRWAGMIASFLLIFGIVAG 785  
  
RESULT 7  
Q99P65 PRELIMINARY; PRT; 475 AA.  
AC Q99P65;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE EQUILIBRATIVE NUCLEOSIDE TRANSPORTER 3.  
GN 4933435C21RIK OR ENT3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL;  
RA Hyde R.J., Cass C.E., Young J.D., Baldwin S.A.;  
RT "The ENT family of eukaryote nucleoside and nucleobase transporters:  
RT recent advances in the investigation of structure/function  
RT relationships and the identification of novel isoforms.";  
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF326986; AAK00957.1; -  
DR MGD; MGI:1918529; 4933435C21RIK.  
DR InterPro; IPR002259; DER_Nucleoside_tran.  
DR Pfam; PF01733; Nucleoside_tran; 1.  
DR PRINTS; PRO1130; DERENTRNSPRT.  
DR ProDom; PD005103; DER_Nucleoside_tran; 1.
```



```

RESULT          9
ID QBGB85
AC QBGB85;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-OCT-2001 (TRENBLrel. 18, Last annotation update)
DE NMPR-INSENSITIVE NUCLEOSIDE TRANSPORTER EI.
GN ENT2.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eute
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RA Wu S.K., Ann D.K., Kim K., Lee V.H.L.;
RT "Molecular cloning and functional characterization of ec
RT nucleic acid transporters (rBENT2 and its constitutively s
RT rBENT2A) in rabbit.";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF323951; AAK11605.1; -.
RL InterPro: IPR002259; DRR_Nucleoside_tran.
DR Pfam: PF01733; Nucleoside_tran; 1.
DR PRINTS: PR01130; DERENTRNSPRT.
DR PRODOM: PD003103; DRR_Nucleoside_tran; 1.
SQ SEQUENCE 456 AA; 49858 MW; 6D55FC9D3D2A392 CRC64;

```







RA Strausberg R. ;  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF131212; AAF78452.1; -  
DR EMBL; AF218255; AAF64036.2; -  
DR EMBL; BC006812; AAH06812.1; -  
DR MGD; MGI:1927073; Slc29a1.  
DR InterPro; IPR002259; DER\_Nucleoside\_tran.  
DR Pfam; PF01733; Nucleoside\_tran; 1.  
DR PRINTS; PR01130; DERENTRNSPRT.  
DR ProDom; PD005103; DER\_Nucleoside\_tran; 1.  
SQ SEQUENCE 458 AA; 50076 MW; EDE0FDEF5BF3E274 CRC64;

Query Match 13.2%; Score 302.5; DB 11; Length 458;  
Best Local Similarity 22.5%; Pred. No. 7.8e-17;  
Matches 104; Conservative 82; Mismatches 215; Indels 61; Gaps 13;

Qy 18 PRDKYNIYWLIVLGVGLPWNMTIAPYEYVNYW-----FKPDGVEYTWYSKEFMGS 72  
Db 7 PDQRYKAWLIFFFVLGTLPLPWNFFMT-ATKYFTNRLDVSNVSSDTQSCESTKALAD 65  
Qy 73 LTIGSQLPNASINVFNLIIAGLPIYVFAVPCFN-----IVNLTIIL 116  
Db 66 PTVALPARSSLSAIFNNVMTLCMLPLLVF--TCLNSPLHORISQSVRIILGSLAILLVF 123  
Qy 117 ILIVIVLEPTEDSMGFWVTLGMATSNFNGLYSENSYVGVGDPPHTYIGALLIGNNIC 176  
Db 124 LVTAALVKVEMDALIFFVITIKIVLINSFGAILQASLFLAGVLPAHYTAPINSQGOLA 183  
Qy 177 GLLITVVKIGVYFLENDEPKLVAIVYFGISLVILLVCAIAFFITKQDFYHHQKGM- 235  
Db 184 GFTTSVAMI-CAIASGSELSESAGFYFTACAVVILAILCYLALPRTFERYHYLQNLAG 242  
Qy 236 -----IRE-----KAETDRPSP-----SILWTFITNCYQLFNVWFCFAVT 271  
Db 243 PAQETKLDLIREPKGRRESGVPGNPSPTNRNQSIKAILKSCVDPAL-SVCFIFTVT 301  
Qy 272 LTIPVMMVTTRGDSGLNKMSENDEIYTLTSFLVFNLFPAAGISIVASKIHW-TP 329  
Db 302 IGLFP---AVTAEVESISAG--TSPWKSIFYIPVACFLNFVFDWLGRLTAVCMWPGQDS 356  
Qy 330 RYLAFAILLRALFIPFFFCNYRQ---TRAYPVFFESTDIFVIGGIAMSFHGYLSALA 386  
Db 357 RNLPLVASRIIVFIPLLMLCNVKARHCGAQRHHFVKHDWFIAFMAAFASNGYLSALC 416  
Qy 387 MGYTPNVVPSHYSRFAQLSVCTLMVGLLTGGLWPVVIEFV 428  
Db 417 MCFGPKVKPAEATAGNIMSFFLCLGLALGAVLSFLRLALV 458

Search completed: July 23, 2002, 16:42:38  
Job time: 496 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 16:35:07 ; Search time 24.68 seconds  
(without alignments)  
680.887 Million cell updates/sec

Title: US-08-816-011f-63  
Perfect score: 2294  
Sequence: 1 MYIINRNTYAVEQEAFFRD.....LTGGLWPVVEHFVDKPSIL 434

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Lasting first 45 summaries

Database : SwissProt\_40

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	343.5	15.0	455	1 ENT1_HUMAN	Q9808 homo sapien
2	326.5	14.2	456	1 ENT2_HUMAN	Q14542 h equilibra
3	320.5	14.0	456	1 ENT2_RAT	O54699 rattus norv
4	301	13.1	456	1 ENT1_RAT	O54698 rattus norv
5	209.5	9.1	327	1 ENT2_MOUSE	Q61672 m equilibra
6	182.5	8.0	517	1 FU26_YEAST	P31381 saccharomyc
7	134	5.8	432	1 V672_METJA	Q58086 methanococc
8	124.5	5.4	506	1 HITB_HAEIN	P1338 haemophilus
9	117.5	5.1	559	1 YAE8_SCHPO	Q9852 schizosacch
10	113.5	4.9	477	1 Y98_MYCGE	P47344 mycoplasma
11	113.5	4.9	587	1 T9S3_MOUSE	Q9et30 mus musculu
12	113.5	4.9	589	1 T9S3_HUMAN	Q9hd45 homo sapien
13	113.5	4.9	604	1 NUSM_HORSE	P48656 equus cabal
14	113	4.9	459	1 TCR2_BACSU	P14512 bacillus su
15	113	4.9	614	1 YDNK_LACIC	P42377 lactococcus
16	112	4.9	383	1 NAPA_ENTHR	P26235 enterococcu
17	111.5	4.9	397	1 TCR2_HUMAN	P55085 homo sapien
18	111.5	4.9	433	1 TCR2_STAAU	P02983 staphylococ
19	111.5	4.9	439	1 UHPC_ECOLI	P09836 escherichia
20	110.5	4.8	479	1 Y98_MYCPN	P75535 mycoplasma
21	110.5	4.8	785	1 ISP4_SCHPO	P40900 schizosacch
22	108.5	4.7	442	1 UHPC_SALTY	P27669 salmonella
23	108	4.7	388	1 YUBA_BACSU	Q32086 bacillus su
24	106.5	4.6	399	1 PAR2_MOUSE	P55086 mus musculu
25	106	4.6	547	1 NUSM_ASCSU	P24884 ascaris suu
26	105.5	4.6	598	1 THIX_YEAST	Q08485 saccharomyc
27	105	4.6	542	1 ATN1_YEAST	P13090 saccharomyc
28	104.5	4.6	894	1 YN86_YEAST	P27514 saccharomyc
29	104	4.5	527	1 NU2M_ACACA	Q37376 acanthameob
30	104	4.5	552	1 NU5M_RHISA	Q9zym7 rhinoceph
31	103.5	4.5	345	1 NUOH_RHOCA	P42032 rhodobacter
32	103.5	4.5	606	1 NU5M_EQUAS	P92485 equus asinu
33	103	4.5	459	1 NU4M_BOVIN	P03910 bos taurus

RESULT	ENT1_HUMAN	1
ID	ENT1_HUMAN	STANDARD; PRT; 455 AA.
AC	Q9808;	
DT	16-OCT-2001 (Rel. 40, Created)	
DT	16-OCT-2001 (Rel. 40, Last sequence update)	
DT	16-OCT-2001 (Rel. 40, Last annotation update)	
DE	Equilibrative nucleoside transporter 1 (Equilibrative nitrobenzylmercaptopyrine riboside-sensitive nucleoside transporter) (Nucleoside transporter, es-type).	
DE	Equilibrative NEMPR-sensitive nucleoside transporter	
GN	SLC29A1 OR ENT1.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 1-21.	
RC	TISSUE=Placenta;	
RC	MEDLINE=97140266; PubMed=8986748;	
RA	Griffiths M., Beaumont N., Yao S.Y.M., Sundaram M., Boumah C.E., Davies A., Kwong F.Y.P., Coe I., Cass C.E., Young J.D., Baldwin S.A.;	
RT	"Cloning of a human nucleoside transporter implicated in the cellular uptake of adenosine and chemotherapeutic drugs."	
RL	Nat. Med. 3:89-93(1997).	
RP	[2]	
RC	SEQUENCE FROM N.A.	
RA	Lum P.Y., Ngo L.Y., Bakken A.H., Unadkat J.D.;	
RT	"Critical structural determinants for high affinity binding of nucleosides to the equilibrative NEMPR-sensitive nucleoside transporter (es) cloned from the human jejunum."	
RL	Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.	
CC	-1- FUNCTION: MEDIATES BOTH INFLUX AND EFFLUX OF NUCLEOSIDES ACROSS THE MEMBRANE (EQUILIBRATIVE TRANSPORTER). IT IS SENSITIVE (ES) TO LOW CONCENTRATIONS OF THE INHIBITOR NITROBENZYLMERCAPTOPURINE RIBOSIDE (NEMPR) AND IS SODIUM-INDEPENDENT. IT HAS A HIGHER AFFINITY FOR ADENOSINE. INHIBITED BY DIPYRIDAMOLE AND DILAZEP (ANTICANCER CHEMOTHERAPEUTICS DRUGS).	
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein.	
CC	-1- TISSUE SPECIFICITY: EXPRESSED IN HEART, BRAIN, MAMMARY GLAND, ERYTHROCYTES AND PLACENTA, AND ALSO IN FETAL LIVER AND SPLEEN.	
CC	-1- PTM: GLYCOSYLATED.	
CC	-1- SIMILARITY: BELONGS TO THE SLC29A FAMILY OF TRANSPORTERS.	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC	-----	
DR	EMBL; U81375; AAC51103.1; -	
DR	EMBL; AF079117; AAC62495.1; -	
DR	MIN; 602193; -	

DR	InterPro; IPR002259; DER_Nucleoside_tran.	AC	Q14542; O43530; Q9UPE0;
DR	Pfam; PF01733; Nucleoside_tran; 1.	DT	01-NOV-1997 (Rel. 35, Created)
DR	PRINTS; PR01130; DERENTRNSPT.	DT	16-OCT-2001 (Rel. 40, Last sequence update)
DR	ProDom; PD005103; DER_Nucleoside_tran; 1.	DT	16-OCT-2001 (Rel. 40, Last annotation update)
KW	Transmembrane; transport; Glycoprotein.	DE	Equilibrative nucleoside transporter 2 (Equilibrative
FT	INIT_MET 0 0	DE	nitrobenzylmercaptopurine riboside-insensitive nucleoside transporter)
FT	DOMAIN 1 11	DE	(Equilibrative NBMPR-insensitive nucleoside transporter) (Nucleoside
FT	TRANSMEM 12 28	DE	transporter, ei-type) (36 kDa nucleolar protein HNP36) (Hydrophobic
FT	DOMAIN 29 81	DE	nucleolar protein; 36 kDa) (Delayed-early response protein 12).
FT	TRANSMEM 82 106	GN	SLC29A2 OR ENT2 OR HNP36 OR DER12.
FT	DOMAIN 107 110	OS	Homo sapiens (Human).
FT	TRANSMEM 111 129	OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
FT	DOMAIN 130 137	OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
FT	TRANSMEM 138 156	OX	NCBI_TaxID=9606;
FT	DOMAIN 157 173	RN	[1]
FT	TRANSMEM 174 198	RP	SEQUENCE FROM N.A.
FT	DOMAIN 199 205	RX	MEDLINE=98148080; PubMed=9478986;
FT	TRANSMEM 206 226	RA	Crawford C.R., Patel D.H., Naeve C., Belt J.A.;
FT	DOMAIN 227 290	RA	"Cloning of the human equilibrative, nitrobenzylmercaptopurine
FT	TRANSMEM 291 310	RT	riboside (NBMPR)-insensitive nucleoside transporter ei by functional
FT	DOMAIN 311 322	RT	expression in a transport-deficient cell line.";
FT	TRANSMEM 323 341	RL	J. Biol. Chem. 273:5288-5293(1998).
FT	DOMAIN 342 358	RN	[2]
FT	TRANSMEM 359 377	RP	SEQUENCE FROM N.A.
FT	DOMAIN 378 392	RC	TISSUE=Placenta;
FT	TRANSMEM 393 412	RX	MEDLINE=98060726; PubMed=9396714;
FT	DOMAIN 413 430	RA	Griffiths M., Yao S.Y., Abidi F., Phillips S.E., Cass C.E.,
FT	TRANSMEM 431 451	RA	Young J.D., Baldwin S.A.;
FT	DOMAIN 452 455	RT	"Molecular cloning and characterization of a nitrobenzylthioinosine-
FT	CARBOHYD 47 47	RT	insensitive (ei) equilibrative nucleoside transporter from human
FT	SEQUENCE 455 AA; 50088 MW; 90989E5E26515850 CRC64;	RL	placenta.";
SQ		RL	Biochem. J. 328:739-743(1997).
		RN	[3]
		RP	SEQUENCE FROM N.A. (SHORT FORM).
		RC	TISSUE=Heart;
		RX	MEDLINE=95367016; PubMed=7639753;
		RA	Williams J.B., Lanahan A.A.;
		RT	"A mammalian delayed-early response gene encodes HNP36, a novel,
		RT	conserved nucleolar protein.";
		RL	Biochem. Biophys. Res. Commun. 213:325-333(1995).
		CC	-1- FUNCTION: MEDIATES EQUILIBRATIVE TRANSPORT OF PURINE, PYRIMIDINE
		CC	NUCLEOSIDES AND THE PURINE BASE HYPOXANTHINE. LESS SENSITIVE THAN
		CC	SLC29A1 TO INHIBITION BY NITROBENZYLTHIOINOSINE (NBMPR),
		CC	DIPYRIDAMOLE, DILAZEP AND DRAFLAZINE.
		CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. HNP36 ISOFORM IS
		CC	A NUCLEOLAR INTEGRAL MEMBRANE PROTEIN.
		CC	-1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
		CC	SHORT FORM/HNP36; SEEMS TO BE PRODUCED BY ALTERNATIVE SPLICING.
		CC	-1- TISSUE SPECIFICITY: EXPRESSED IN PLACENTA, BRAIN, HEART AND
		CC	OVARIAN TISSUES.
		CC	-1- INDUCTION: BY PLATELET DERIVED GROWTH FACTOR (PDGF) AND FIBROBLAST
		CC	GROWTH FACTOR (FGF).
		CC	-1- SIMILARITY: BELONGS TO THE SLC29A FAMILY OF TRANSPORTERS.
		CC	-----
		CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
		CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
		CC	the European Bioinformatics Institute. There are no restrictions on its
		CC	use by non-profit institutions as long as its content is in no way
		CC	modified and this statement is not removed. Usage by and for commercial
		CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/
		CC	or send an email to license@isb-sib.ch).
		CC	-----
		CC	EMBL; AF034102; AAB97834.1; -
		DR	EMBL; AF029358; AAC39526.1; -
		DR	EMBL; X86681; CAA60380.1; -
		DR	MM; 602110; -
		DR	InterPro; IPR002259; DER_Nucleoside_tran.
		DR	Pfam; PF01733; Nucleoside_tran; 1.
		DR	PRINTS; PR01130; DERENTRNSPT.
		DR	ProDom; PD005103; DER_Nucleoside_tran; 1.
		DR	Nuclear protein; Transmembrane; Transport; Alternative splicing.
		FT	TRANSMEM 13 33
		FT	TRANSMEM 70 90
		FT	POTENTIAL.
		FT	POTENTIAL.

Query Match	15.0%;	Score	343.5;	DB	1;	Length	455;
Best Local Similarity	24.3%;	Pred. No.	1.2e-16;				
Matches 112; Conservative	77;	Mismatches	212;	Indels	59;	Gaps	13;
QY	18	PRDKYNIYVWLVGFGVLLPWNFFITTAPEYVYVWYVFKPDGVE--TWYSKEPMGSLT	74				
Db	6	PODRYKAVWLFFMGLGTLTPWFFMT-ATQYFNRLDMSQNSLVTAELSKDAQAA	64				
QY	75	IGSQLP--NASINVNFLIITAGPLIYRVA-----PVCNIVN-----ITILI	117				
Db	65	PAAPLPERNLSAIFNNVMTLCAMPLPLLTLYLNSFLHQRIPQSVRIILGSLVAILLVFLI	124				
QY	118	LVIVLEPTEDSMSPFWTTLGMATSNFNSGLYNSVVGSGDRPHTVIGALLIGNNICG	177				
Db	125	TAILVKVQLDALP-FFVITMKIVLINSEFALQGLSLGLAGLLPASTAPIMSQGLAG	183				
QY	178	LLITVVKIGVTFYFLNDEPKLVAIVFGISLVILVCAIALFFITKQDFYHHQ-----	231				
Db	184	FFASVAMI-CAIAGSSELSAFAFYITACAVIILTIICYLGLPRLEFYRYQQKLBSGP	242				
QY	232	-----KGMEIRKAETDRSPSILMT-----FNCYQQLFNWFCFVAVT	271				
Db	243	GEQETKLDLSKGEPRAGKEESGVSVNSQPTNESHSAIKAILKNISVLAFSVCFITFT	302				
QY	272	LTIFPVMVTITRGDSGLFNKIMSENDEIYTLTSLFVNLFAAIGSIVASKIHWP--TP	329				
Db	303	IGMPFPAVTEVKSSTAG-----SSTWEYFIPVSCFLTFNFDWIGRLSLTAVFWMPGKDS	357				
QY	330	RYLKFAIILRALFIPIFFFCNRYVQTRAY-PVFFESTDIFVIGGIAMSFHGYLSALAMG	388				
Db	358	RWPLSLVLARLVFVPLLCLN--IKPRYLIVVFEHDAWFFIFFMAAFASNGYLASLCMC	415				
QY	389	YTPNVVPSHYGRFAQLSVCTLMVGLLGGGLMPVVVIEHV	428				
Db	416	FGPKVKPAEATAGAIMAFFLCGLGALGAVSFLEIRAIV	455				
RESULT	2						
ENT2_HUMAN							
ID	ENT2_HUMAN	STANDARD;	PRT;	456	AA.		



```
FT TRANSMEM 99 119 POTENTIAL.
FT TRANSMEM 124 144 POTENTIAL.
FT TRANSMEM 162 182 POTENTIAL.
FT TRANSMEM 193 213 POTENTIAL.
FT TRANSMEM 291 311 POTENTIAL.
FT TRANSMEM 324 344 POTENTIAL.
FT TRANSMEM 360 380 POTENTIAL.
FT TRANSMEM 386 406 POTENTIAL.
FT TRANSMEM 432 452 POTENTIAL.
FT CARBOHYD 48 48 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 225 225 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 1 130 MISSING (IN ISOFORM HNP36).
FT CONFLICT 200 200 Y -> C (IN REF. 1).
SQ SEQUENCE 456 AA; 50172 MW; AEF1B244397508E1 CRC64;

Query Match 14.2%; Score 326.5; DB 1; Length 456;
Best Local Similarity 22.9%; Pred. No. 1.7e-15;
Matches 106; Conservative 71; Mismatches 200; Indels 85; Gaps 11;

QY 18 PRDXYNIYVWLVILVGVGLLPNNFFITIAPEYY-----VNYWEKPDGVE 62
D 7 PRDXYNIYVWLVILVGVGLLPNNFFITIAPEYYQARAGANSTARILSTNHTGPDFAF- 65
QY 63 TWYSKEFGSLTIGSQPNASINVFNLFIITAGPLIYRVFAPVCNFIIVNLTIILIVL 122
D 66 -----NENWVTLSQLPPLLFTLLNSFLYQCVETVRLG-----SLAILLFTALTAAL 116
QY 123 EPTEDSWFVWTLGMATSNFNGLYENSYGVGGDPHYTGALLIGNICGLLTV 182
D 117 VKVDMSPGPFFSTWASVCNFSNFSVQLGSLGQTMPTSTYTLFSLGGGLAGIFAA 176
QY 183 VKI-----GVTYFLNDEPKLVAIYF---GISVLIVLCAATLFIKQDPY---HYHQ 231
D 177 AMLLSMASGV-----DAETSALGYFTIPYVGLMSIVCYLSLPHLFARYLANKSSQA 230
QY 232 KGBIREKA-----ETDRPSPSILWTTFNC 257
D 231 QAQELTKAELQSDENGIPSPQKVALTLDLLEKPESEPDPEPQKGRKSVFTVQKI 290
QY 258 YGQLENWFCAVTLTTPVMMVTTRGDSFLNKNINSENDIYTLTSLFVFNLFRAIG 317
D 291 WLTAALCLVLTVTLVFPFAITAVTSTST-----PGKWSQFFNPICCFLLNFMWL 344
QY 318 SIVASKIHP--TPRYLKFAILRALFIPFFFCNYRVOTRAYVFEESTDIFVIGGIA 375
D 345 RSLTSYFLWPDDESRLLPLVLCRLFLVPLFMCHVFORSR-LPILFPQDAYFITFMLLF 403
QY 376 SFSHGYSALAMGYTPNVVPSHFRFAQLSVCTLMVGLLTG 417
D 404 AVSNGYLVSLTMCLAPQVLPHEREVAGALMTFFFLALGLSCG 445

RESULT 3
ENT2_RAT STANDARD; PRT; 456 AA.
AC OS4699;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Equilibrative nucleoside transporter 2 (Equilibrative
DE nitrobenzylmercaptopyrine riboside-insensitive nucleoside transporter)
DE (Equilibrative NBMPR-insensitive nucleoside transporter) (Nucleoside
DE transporter, ei-type).
GN SLC29A2 OR ENT2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Jejunum;
```

```
RX MEDLINE=98019212; PubMed=9353301;
RA Yao S.Y.M., Ng A.M.L., Muzyka W.R., Griffiths M., Cass C.E.,
RA Baldwin S.A., Young J.D.;
RT "Molecular cloning and functional characterization of
RT nitrobenzylthioinosine (NBMPR)-sensitive (es) and NBMPR-insensitive
RT (ei) equilibrative nucleoside transporter proteins (rENT1 and rENT2)
RT from rat tissues.";
RL J. Biol. Chem. 272:28423-28430(1997).
CC -1- FUNCTION: MEDIATES BOTH INFUX AND EFFLUX OF NUCLEOSIDES ACROSS
CC THE MEMBRANE (EQUILIBRATIVE TRANSPORTER). IT IS INSENSITIVE (EI)
CC TO LOW CONCENTRATIONS OF THE INHIBITOR NITROBENZYLMERCAPTOPURINE
CC RIBOSIDE (NBMPR) AND IS SODIUM-INDEPENDENT. SPECIFIC FOR
CC NUCLEOSIDES, BUT MAY ALSO TRANSPORT HYPOXANTHINE. MAY ALSO PLAY A
CC ROLE IN THE EFFLUX OF INOSINE AND HYPOXANTHINE FROM MUSCLE CELLS
CC DURING THE NET DEGRADATION OF PURINE NUCLEOTIDES THAT OCCURS
CC DURING STRENUOUS EXERCISE AND/OR IN THE REUPTAKE OF THESE PURINES
CC DURING THE RECOVERY PROCESS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- MISCELLANEOUS: RESISTANT TO DIPYRIDAMOLE AND DILAZEP INHIBITION
CC (ANTICANCER CHEMOTHERAPEUTICS DRUGS).
CC -1- SIMILARITY: BELONGS TO THE SLC29A FAMILY OF TRANSPORTERS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF015305; AAB88050.1;
CC InterPro: IPR002259; DER_Nucleoside_tran.
CC Pfam: PF01733; Nucleoside_tran; 1.
CC PRINTS: PR01130; DERENTRNSPRT.
CC ProDom: PD005103; DER_Nucleoside_tran; 1.
CC Transmembrane; Transport; Glycoprotein.
CC DOMAIN 1 12 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 13 29 POTENTIAL.
CC DOMAIN 30 68 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 69 93 POTENTIAL.
CC DOMAIN 94 97 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 98 116 POTENTIAL.
CC DOMAIN 117 124 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 125 143 POTENTIAL.
CC DOMAIN 144 160 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 161 185 POTENTIAL.
CC DOMAIN 186 192 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 193 213 POTENTIAL.
CC DOMAIN 214 291 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 292 311 POTENTIAL.
CC DOMAIN 312 323 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 324 342 POTENTIAL.
CC DOMAIN 343 359 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 360 378 POTENTIAL.
CC DOMAIN 379 393 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 394 413 POTENTIAL.
CC DOMAIN 414 431 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 432 452 POTENTIAL.
CC DOMAIN 453 456 EXTRACELLULAR (POTENTIAL).
CC TRANSMEM 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC CARBOHYD 56 56 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 456 AA; 50265 MW; DA97C2C578E1EE9D CRC64;

Query Match 14.0%; Score 320.5; DB 1; Length 456;
Best Local Similarity 23.5%; Pred. No. 4.4e-15;
Matches 107; Conservative 76; Mismatches 200; Indels 73; Gaps 13;

QY 18 PRDXYNIYVWLVILVGVGLLPNNFFITIAPEYYNIYVNFKPDGVETWYSK-----EF 69
D 7 PRDXYNIYVWLVILVGVGLLPNNFFITIAPEYYNIYVNFKPDGVETWYSK-----EF 69
QY 70 MGSLTIGSQLPNASINVFNLFIITAGPLIYRVFAPVCNFIIVNLTIILIVLVIPTEDSM 129
```

Db 67 NNVVTLTLLSPLTLLNSFLYQIPESVRLG-----SLAILLFALTAALVKVDLSP 122  
Qy 130 SFFFWTLTGMATSNFNSGLVNSVYGGDPHTYIGALLIGNNICGL-----LITVVK 184  
Db 133 GLFFSITWASVWFNSFCVAVLGSLGQGLTWPSTYSLFLSGQGLAGTFAALAMLTSLA 182  
Qy 185 IGVTYFLNDEPKLVAIVF---GISLVILLVCAIAL-----FFITKQ----- 223  
Db 183 SGV-----DQTSALGVFITPCVGILLISICYLSPHLKFARYLYTKKQPAPVQLETK 236  
Qy 224 -DFYHYHKGMEIR-----EK-----AETDRSPSILWTFNFCYQQLNV 264  
Db 237 AELLGADEKNGIPVSPQAGPTLDLDPKELEGLGELGEPKPKPSVFFVFRKIMLTALCL 296  
Qy 265 WFCFAVTLTIPP-VMMVTYTRGDSGLFNKINSEDEIYTLTSLFVNLFAAIGSVASK 323  
Db 297 VLVFVTLVSVPFAITAMVTTSSN-----PKWSQFFNPICCFLLFFNMDWLGRLSY 350  
Qy 324 IHWP---TPRYLKFAILLRALFIPFFFCNRYVQTRAYVFFESTDIFVIGGTAMSFSGY 381  
Db 351 FLWPEDESQLPLVLCFLFVPLFMLCHVQPAR-LPIFWQDAYFITEMLLFAISNGY 409  
Qy 382 LSLAMGTTPNVVPSHYSRFAQLSVCTLMVGLLTG 417  
Db 410 FVSLTMCPLAQVLPHEREVAGALMTFFFLALGLSCG 445

RESULT 4  
ENTL\_RAT STANDARD; PRT; 456 AA.  
ID ENTL\_RAT AC  
OS 054698;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Equilibrative nucleoside transporter 1 (Equilibrative  
nitrobenzylmercaptopyrimidine riboside-sensitive nucleoside transporter)  
DE (Equilibrative NBMPR-sensitive nucleoside transporter) (Nucleoside  
transporter, es-type).  
GN SLC29A1 OR ENT1.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_Taxid=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Jejunum;  
RX MEDLINE=98019212; PubMed=9353301;  
RA Yao S.Y.M., Ng A.M.B., Muzika W.R., Griffiths M., Cass C.E.,  
RA Baldwin S.A., Young J.D.;  
RT "Molecular cloning and functional characterization of  
nitrobenzylthioinosine (NBMPR)-sensitive (es) and NBMPR-insensitive  
(ei) equilibrative nucleoside transporter proteins (rENT1 and rENT2)  
from rat tissues.";  
RL J. Biol. Chem. 272:28423-28430(1997).  
CC -!- FUNCTION: MEDIATES BOTH INFLUX AND EFFLUX OF NUCLEOSIDES ACROSS  
THE MEMBRANE (EQUILIBRATIVE TRANSPORTER). IT IS SENSITIVE (ES) TO  
LOW CONCENTRATIONS OF THE INHIBITOR NITROBENZYLMECAPTOPURINE  
RIBOSIDE (NBMPR) AND IS SODIUM-INDEPENDENT. IT HAS A HIGHER  
AFFINITY FOR ADENOSINE AND IS SODIUM-INDEPENDENT. IT HAS A HIGHER  
INHIBITION (ANTICANCER CHEMOTHERAPEUTICS DRUGS).  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- TISSUE SPECIFICITY: EXPRESSED IN JEJENUM, LIVER AND LUNG.  
CC -!- PTM: GLYCOSYLATED.  
CC -!- SIMILARITY: BELONGS TO THE SLC29A FAMILY OF TRANSPORTERS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AF015304; AAB88049.1; -  
DR InterPro; IPR002259; DER\_Nucleoside\_tran.  
DR Pfam; PF01733; Nucleoside\_tran; 1.  
DR PRINTS; PR01130; DERENTRNSPRT.  
DR PRODOM; PD003103; DER\_Nucleoside\_tran; 1.  
KW Transmembrane; Transport; Glycoprotein.  
FT INT\_MET 0 0  
FT DOMAIN 1 11 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 12 28 POTENTIAL.  
FT DOMAIN 29 81 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 82 106 POTENTIAL.  
FT DOMAIN 107 110 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 111 129 POTENTIAL.  
FT DOMAIN 130 137 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 138 156 POTENTIAL.  
FT DOMAIN 157 173 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 174 198 POTENTIAL.  
FT DOMAIN 199 205 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 206 226 POTENTIAL.  
FT DOMAIN 227 290 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 291 310 EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 311 322 POTENTIAL.  
FT TRANSMEM 323 342 POTENTIAL.  
FT DOMAIN 343 359 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 360 378 POTENTIAL.  
FT DOMAIN 379 393 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 394 413 POTENTIAL.  
FT DOMAIN 414 431 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 432 452 POTENTIAL.  
FT DOMAIN 453 456 EXTRACELLULAR (POTENTIAL).  
FT CARBOHYD 43 43 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 47 47 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 53 53 N-LINKED (GLCNAC. .) (POTENTIAL).  
SQ SEQUENCE 456 AA; 49885 MW; A34CE92C20836D9B CRC64;

Query Match 13.1%; Score 301; DB 1; Length 456;  
Best Local Similarity 22.3%; Pred. No. 9.3e-14;  
Matches 107; Conservative 84; Mismatches 191; Indels 98; Gaps 17;

Qy 18 PRDKNYIVWLVLVGVGVLLPWNMTTIAPEYVYVWFKPDGV-----ETWYSKEFGMS 72  
Db 6 PQDRTKAVMLFFLGLGLTLLPWNFFIT-ATQYETSLRNTSQNISLVTNQSCSEALAD 64  
Qy 73 LTIGSQLPNASINVENFLIITAGPLIYRVEAPVCFN-----IVNLITII 115  
Db 65 PSVSLPARSSLSAIFNNVTYCAMLPDLIF--TCNLSFLHKVKSLSRLGLSLAILLVF 122  
Qy 116 LILVIVLEPTEDSMSEWFFVTGLMATSNFNSGLYENSVYGVGGDPHTYIGALLIGNNI 175  
Db 123 LVTATLVKQMDALS-FFIITMIKVLINSFGAILQASLFGLAGVLPANYTAPIMSGOGL 181  
Qy 176 CGLIITVVKI-----GWTFYFLNDEPKLVAIVYFGISLVILLVCAIALEFFITKQ 223  
Db 182 AGFTSVAMICAVASGSKLSSEAFQFIT---ACAVV-----ILAILCYLALPWW--- 228  
Qy 224 DFYHYHHQ-----KQMEIR-EKAEITDRPSPIL----- 250  
Db 229 EFRHYQLNLNAGPAEQTKLDLISEGEEPRGREGSVPGPSNPANRNSIKAILKSI 288  
Qy 251 WTTFTNCGQLFVWFCFPAVTLTIFFPVMVTYTRGDSGLFNKINSEDEIYTLTSLFVLF 310  
Db 289 WVL-----ALSVCFTFTVIGLFP---AVTAEVESIACTSPWKNK-YEIPVACFLNF 337  
Qy 311 NLFRAIGSIVASKIHWP---TPRYLKFAILLRALFIPFFFCNRYVQTRAYVFFESTDIF 368  
Db 338 NVFDWLGRLTAICMWPQDSRWLPVLVACRVVFIPLMLCNVK-OHHYLPSELFKHDWVF 396  
Qy 369 VIGGTAMSFSGYLSALAMGTTPNVVPSHYSRFAQLSVCTLMVGLLTGGLWLPVVIEHFV 428  
Db 397 ITFMAAFASNGYLSLASCMECPGPKKVPAPAEATAGNIMSFFLCLGLGALGVLFLRLALV 456

```

RESULT 5
ENT2_MOUSE STANDARD; PRT; 327 AA.
AC O61672;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Equilibrative nucleoside transporter 2 (Equilibrative
DE Nitrobenzylmercaptopurine riboside-insensitive nucleoside transporter) (Nucleoside
DE (Equilibrative NEMPR-insensitive nucleoside transporter) (Nucleoside
DE transporter, ei-type) (36 kDa nucleolar protein HNP36) (Hydrophobic
DE nucleolar protein, 36 kDa) (Delayed-early response protein 12).
DE SLC29A2 OR ENT2 OR HNP36 OR DER12.
OS Mus musculus (Mouse).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Fibroblast;
RX MEDLINE=95367016; PubMed=7639753;
RA Williams J.B., Lanahan A.A.;
RT "A mammalian delayed-early response gene encodes HNP36, a novel,
RT conserved nucleolar protein.";
RL Biochem. Biophys. Res. Commun. 213:325-333(1995).
CC -1- FUNCTION: MEDIATES EQUILIBRATIVE TRANSPORT OF PURINE AND
CC PYRIMIDINE NUCLEOSIDES, AND THE PURINE BASE HYPOXANTHINE.
CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLAR. INTEGRAL MEMBRANE
CC PROTEIN
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM AND A SHORT
CC FORM/HNP36 (SHOWN HERE); SEEMS TO BE PRODUCED BY ALTERNATIVE
CC SPLICING.
CC -1- INDUCTION: BY PLATELET DERIVED GROWTH FACTOR (PDGF) AND FIBROBLAST
CC GROWTH FACTOR (FGF).
CC -1- SIMILARITY: BELONGS TO THE SLC29A FAMILY OF TRANSPORTERS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X86682; CAA60381.1;
CC MGD; MGI:1345278; SLC29a2.
DR InterPro: IPR002259; DER_Nucleoside_tran.
DR Pfam; PF01733; Nucleoside_tran; 1.
DR PRINTS; PR01130; DERENTRNSPRT.
DR PRODOM; PD005103; DER_Nucleoside_tran; 1.
KW Nuclear protein; Transmembrane; Transport; Alternative splicing.
FT TRANSMEM 2 22 POTENTIAL.
FT TRANSMEM 32 52 POTENTIAL.
FT TRANSMEM 63 83 POTENTIAL.
FT TRANSMEM 159 179 POTENTIAL.
FT TRANSMEM 194 214 POTENTIAL.
FT TRANSMEM 231 251 POTENTIAL.
FT TRANSMEM 267 287 POTENTIAL.
FT TRANSMEM 303 323 POTENTIAL.
SQ SEQUENCE 327 AA; 36113 MW; 5D2D3FF4BBD59286 CRC64;

Query Match 9.1%; Score 209.5; DB 1; Length 327;
Best Local Similarity 22.4%; Pred. No. 11e-07;
Matches 76; Conservative 49; Mismatches 134; Indels 81; Gaps 12;

Qy 131 WFFWYTLGMATSFINSGLYENSYGVGGDPFPHYTGALLGNICGLLITVVKI----- 185
|| : : : : : || : : : : :
Db 5 WF-----INSFCVGLSLFGGLGTWPTSTYTLFSLSGQGLAGIFAALMLSLAS 54
Qy 186 GVTYFLNDEPKLVAIVF---GISVLVLLVCAIAL-----FTI-----TK 222
|| : : : : : || : : : : :

```

```

Db 55 GV-----DAQTSALGYFITPCVGILLISIVCYLSLPHLKFAFYLYTEKLSQAQPTOLETK 108
Qy 223 QDFVYHHQKMEIR-----EKAETDRSPSILWTTFTNCYQGLFNWVF 266
Db 109 AELQADEKNGVPISPQOASPTLDLDPEKEPEPEPQKPSVFWFRKIWLTAICLVL 168
Qy 267 CFAVLTITFPVM--MTVTRGDSG----FLNKIMSENDEIYTLTSLFVFNLFAAIGSIV 320
Db 169 VFTVTLVSFPAITAMVTTSSNSPGKWLFFNP1-----CCFLLFNMDMLGRSL 217
Qy 321 ASKIHWP---TPRYLKFAIILRALPIPFPPFCNRYRQVTRAPVVFESTDIFVIGIAMS 377
Db 218 TSYFLWPEDESQLPLLEVLCLRFLEVPFLMCLCHVPOHAR-LPIIFRODAYFITFMLLFAV 276
Qy 378 SHGYLSALAMGYTPNVVPSHYSRFAAQLSVCTLMVGLLTG 417
Db 277 SNGYLVSLTMCLAPROVLPHEREVAGALMTFFLAUGLSCG 316

RESULT 6
FU26_YEAST STANDARD; PRT; 517 AA.
ID FU26_YEAST
AC P31381;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nucleoside transporter FUN26.
GN FUN26 OR YAL022C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=93209532; PubMed=8458570;
RA Ouellette B.F., Clark M.W., Keng T., Storms R.K., Zhong W.W.,
RA Zeng B., Fortin N., Delaney S., Barton A.B., Kaback D.B., Bussey H.;
RT "Sequencing of chromosome I from Saccharomyces cerevisiae: analysis
RT of a 32 kb region between the LTEL and SPO7 genes.";
RL Genome 36:32-42(1993).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=20408929; PubMed=10827169;
RA Vickers M.F., Yao S.Y., Baldwin S.A., Young J.D., Cass C.E.;
RT "Nucleoside transporter proteins of Saccharomyces cerevisiae.
RT Demonstration of a transporter (FUI1) with high uridine selectivity
RT in plasma membranes and a transporter (FUN26) with broad nucleoside
RT selectivity in intracellular membranes.";
RL J. Biol. Chem. 275:25931-25938(2000).
CC -1- FUNCTION: HAS BROAD NUCLEOSIDE SELECTIVITY (URIDINE, ADENOSINE AND
CC CYTIDINE) AND MOST LIKELY FUNCTIONS TO TRANSPORT NUCLEOSIDES
CC ACROSS INTRACELLULAR MEMBRANES.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE SLC29A FAMILY OF TRANSPORTERS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L05146; AAC04935.1;
CC DR EMBL; L05027; AAA70393.1;
CC DR PIR; S36712; S36712.
CC DR SGD; S0000020; FUN26.
CC DR InterPro: IPR002259; DER_Nucleoside_tran.
CC DR Pfam; PF01733; Nucleoside_tran; 1.
CC DR PRODOM; PD005103; DER_Nucleoside_tran; 1.
KW Transmembrane; Transport.
FT TRANSMEM 76 96 POTENTIAL.

```



FT	VARIANT	104		G -> C (IN STRAIN TN106).
FT	VARIANT	116		V -> G (IN STRAIN TN106).
FT	VARIANT	167		S -> F (IN STRAIN TN106).
FT	VARIANT	252		T -> M (IN STRAIN TN106).
SQ	SEQUENCE	506 AA;	56546 MW;	ADA28861C1481AID CRC64;
Query Match            5.4%; Score 124.5; DB 1; Length 506; Best Local Similarity 21.6%; Pred. No. 0.11; Matches 105; Conservative 69; Mismatches 144; Indels 167				
QY	26 YW---	LVLIVGCVLLPWNMFITIAPEY-----YYNWFKPDGVETWSKE-FMGSLL	:	: : : :
DB	7 FWLTLLIILIGLPLCPFLFYVLRATEVGTRSVLELFRRPMAEILLSNTMLLMVCV	:	: : : :	:
QY	78 QLPNASINVFNLIAGPLVIYFAPCFNVNLTLIIILVILEPTEDSMW----	:	: : : :	:
DB	67 ----ISLGTFCAELLER----YRFFGKAFEEVAMTLCIPAFVSGFTWISLTRFRV	:	: : : :	:
QY	135 VTGLMATSFNFI-----SNGLYENSIVGYGDPPHYTGAL-----L	:	: : : :	:
DB	119 GTIGIMTLSSFPPLAVLPVSAILKRDLRSLEEVSLSLKSPVYFWVAISPQLKPA	:	: : : :	:
QY	175 ICGLLIT---VVKGIVTYEFLNDPKLVAI-----VYEG-----ISLVILLVCAI-	:	: : : :	:
DB	178 L--LLIALHMLEVFEGAVSLNYQTFTAIFQEYESFNNSSTAALLSAVMAICILI	:	: : : :	:
QY	217 LFFITKDQFYHHVKHGMEIREKAETDRSPSILTWTFTNCYGQ-----LFNWWF	:	: : : :	:
DB	236 IFFRGQTL--YHSKGKV-----TRP-----YLKTLSFGKOCLTFGFESSIF	:	: : : :	:
QY	271 TLTIFFPVMM-----TWTTRGD-----SGFLNK-IMSENDRIYTLLTSF-	:	: : : :	:
DB	281 GV---PVMILYWLIVGTLESAGDFSEFLSAFSNFIIISGLGALLTVMCALPLVWM	:	: : : :	:
QY	308 -----LVFNLFQAIG-----SI VASKITHWPTRYKLFAILLRALFI-	:	: : : :	:
DB	338 YRSYLTIWIRLPIYLHVAPGLVIALSLVFSIHANDLYQTFFVIIITAYEMXLYLP	:	: : : :	:
QY	344 -----PERFEFCYNRYQTRAYPVFFESTDIFVIGGIAMS	:	: : : :	:
DB	398 TLRASLEQLSDQIEKVGQSGRNPXYIF-----RTLTPA-----ILPGVAAN	:	: : : :	:
QY	381 YLSAL 385	:	: : : :	:
DB	445 FLNLM 449	:	: : : :	:
RESULT 9 YAEC_SCHPO STANDARD; PRT; 559 AA. AC Q09852; DT 01-FEB-1996 (Rel. 33, Created) DT 01-FEB-1996 (Rel. 33, Last sequence update) DE 01-MAR-2002 (Rel. 41, Last annotation update) DE Putative inorganic phosphate transporter C23D3.12. GN SPAC23D3.12. OS Schizosaccharomyces pombe (Fission yeast). OC Eukaryota; Fungi; Ascomycota; Schizosaccaromycetes; OC Schizosaccaromycetales; Schizosaccaromycetaceae; OC Schizosaccaromycetes. OX NCBI_TaxID=4896; RN [1] RP SEQUENCE FROM N.A. RC STRAIN=972; RA Niblett D.; Harris D., Barrell B.G., Rajandream M.A., Walsh S. RL Submitted (OCT-1995) to the EMBL/GenBank/DDBJ databases. CC !- FUNCTION: HIGH-AFFINITY TRANSPORTER FOR EXTERNAL INORGANI CC PHOSPHATE (BY SIMILARITY). CC !- SUBCELLULAR LOCATION: Integral membrane protein (Probable CC !- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY. STRO CC YEAST PHOB4.				

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: Z64354; CAA91247.1; -  
InterPro: IPR003662; sub\_transporter.  
DR Pfam: PF00083; sugar\_tr; 1.  
DR PROSITE: PS00216; SUGAR\_TRANSPORT\_1; 1.  
DR PROSITE: PS00217; SUGAR\_TRANSPORT\_2; 1.  
KW Hypothetical protein; Phosphate transport; Transmembrane.  
FT DOMAIN 1 45 CYTOPLASMIC (POTENTIAL).  
FT TRANSSEM 46 66 1 (POTENTIAL).  
FT DOMAIN 67 94 EXTRACELLULAR (POTENTIAL).  
FT TRANSSEM 95 115 2 (POTENTIAL).  
FT DOMAIN 116 118 CYTOPLASMIC (POTENTIAL).  
FT TRANSSEM 119 139 3 (POTENTIAL).  
FT DOMAIN 140 144 EXTRACELLULAR (POTENTIAL).  
FT TRANSSEM 145 165 4 (POTENTIAL).  
FT DOMAIN 166 188 CYTOPLASMIC (POTENTIAL).  
FT TRANSSEM 189 209 5 (POTENTIAL).  
FT DOMAIN 210 224 EXTRACELLULAR (POTENTIAL).  
FT TRANSSEM 225 245 6 (POTENTIAL).  
FT DOMAIN 246 335 CYTOPLASMIC (POTENTIAL).  
FT TRANSSEM 336 356 7 (POTENTIAL).  
FT DOMAIN 357 381 EXTRACELLULAR (POTENTIAL).  
FT TRANSSEM 382 402 8 (POTENTIAL).  
FT DOMAIN 403 409 CYTOPLASMIC (POTENTIAL).  
FT TRANSSEM 410 430 9 (POTENTIAL).  
FT DOMAIN 431 438 EXTRACELLULAR (POTENTIAL).  
FT TRANSSEM 439 459 10 (POTENTIAL).  
FT DOMAIN 460 474 CYTOPLASMIC (POTENTIAL).  
FT TRANSSEM 475 495 11 (POTENTIAL).  
FT DOMAIN 496 497 EXTRACELLULAR (POTENTIAL).  
FT TRANSSEM 498 518 12 (POTENTIAL).  
FT DOMAIN 519 559 CYTOPLASMIC (POTENTIAL).  
FT TRANSSEM 559 AA; 62316 MW; 2E8958F86C2092E2 CRC64;  
SEQUENCE 559 AA; 62316 MW; 2E8958F86C2092E2 CRC64;  
Query Match 5.1%; Score 117.5; DB 1; Length 559;  
Best Local Similarity 20.1%; Pred. No. 0.35;  
Matches 101; Conservative 73; Mismatches 175; Indels 153; Gaps 25;  
QY 26 YWLV-----LVFG-----VLLPNNFI--TIAPEYVNVWFKPDGVETWYSKFMGSL 73  
DB 34 HWLGLTKREFKLMGFAGFFLDYDLFIINLVSPIVEYLYWGLEKKPKHPYSGIHG-- 91  
QY 74 TIGSQLPNASINVENFLIILAGPLIYR-----VFAPVCNIVNLTILILVIVLE 123  
DB 92 -----LYNAANIGNVF-----GQILFGMGDFGKFKFYGKEMIVVIATVILALPKSI 142  
QY 124 PTE-DSMSWFF--WVTGLMATSNFNSGLYENSVYGVGGDFPHYI-----GA 168  
DB 143 PTLPLKMWIFARWL-LGL-----GIGGDYPMSTATISERSLLSRGT 185  
QY 169 LLI-----GNCCGLLITVWKIGVYFLNDEP-----KLVAI--VYFGSLVILL 211  
DB 186 LLSIVSFQSGFTLAGAIVTILLACF-----EKPLNORGEYTKLEGVWRQLQMLALVPAL 241  
QY 212 VCAI-----ALFFITKDFY-----HYHHKGMIREKAET-- 242  
DB 242 LVLLPLRTMKESYEQSKALKNKTTDNDTYIADDEPKKDNQNVVEKQINLTSSSGHP 301  
QY 243 -----DRSPSILWTFITNCYQQLFNWPCF-----AVTLTIFPVMFTVTRGDSGF 289  
DB 302 TSTEDFGDKRASTVPTSGTSGFIEYSQWHFKHLLATAVSWFLDIAFGVNLQSVI 361  
QY 290 LNKI--HSENDIEVTLTSELVFNLFPAIGSIVASKTHWPT-----PRYLKFAIILRALFI 343  
DB 362 LKAIFGSSGKNEYHTLMRG-AIGNLNTAIAAGYVPG--YWTFVFLVEKLGKRWIQLQGLFI 418

QY 344 PFREFCNRVQTRAYPVFFESTDIEFVIGG-----IAMSFSHGYSALAMGYTPNVVPSH 397  
DB 419 TGLMFA-----ILAGSWDTTSTGRTACFVIAOFFSNGFNATFLYPAEVFFAR 468  
QY 398 YSRFAAQISVCTLMVGLTGGGL 419  
DB 459 VRGTAHGLSALGKCGAILASL 490  
RESULT 10  
Y098\_MYCGE  
ID Y098\_MYCGE STANDARD; PRT; 477 AA.  
AC P47344; Q49231; Q49509;  
DC 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical protein MG098.  
GN MG098.  
OS Mycoplasma genitalium.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;  
OC Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=2097;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 33530 / G-37;  
RX MEDLINE=96026346; PubMed=7569993;  
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,  
RA Fleischmann J.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,  
RA Fritchmann J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.D.,  
RA Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,  
RA Tomb J.-F., Dougherty B.A., Bost K.F., Hu P.-C., Lucier T.S.,  
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.,  
RA "The minimal gene complement of Mycoplasma genitalium.";  
RT Science 270:397-403(1995).  
RN [2]  
RP SEQUENCE OF 465-155 AND 278-382 FROM N.A.  
RC STRAIN=ATCC 33530 / G-37;  
RX MEDLINE=94075230; PubMed=8253680;  
RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;  
RT "A survey of the Mycoplasma genitalium genome by using random  
sequencing.";  
RT J. Bacteriol. 175:7918-7930(1993).  
RL -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).  
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE GATC FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: U39690; AAC71316.1; -  
DR EMBL: U01701; AAB01013.1; -  
DR EMBL: U01782; AAD12771.1; -  
DR TIGR: MG098; -  
DR InterPro: IPR003837; Glu-trnAGln.  
DR Pfam: PF02686; Glu-trnAGln; 1.  
KW Hypothetical protein; Transmembrane; Complete proteome.  
FT TRANSSEM 23 43 POTENTIAL.  
FT TRANSSEM 61 81 POTENTIAL.  
FT TRANSSEM 108 128 POTENTIAL.  
FT TRANSSEM 131 151 POTENTIAL.  
FT TRANSSEM 171 191 POTENTIAL.  
FT TRANSSEM 228 248 POTENTIAL.  
FT TRANSSEM 285 305 POTENTIAL.  
FT TRANSSEM 326 346 POTENTIAL.  
FT CONFLICT 87 F->V (IN REF. 2).  
FT CONFLICT 372 382 DMOGLIMPVNI -> EYAKLNYAOY (IN REF. 2).  
FT SEQUENCE 477 AA; 54045 MW; 3FCCDD95A2C35684 CRC64;









```
QY 151 -----ENSYGVGGDPHTYIGALLIGN-----NIC-GLLTIVVKVIGTYFLNDEP 195
Db 238 PTPVSALLHSTMVAGVFLIRFPLMENNKTIOSLTCLGAIPTLTAICALTQNDIK 297
QY 196 KLVA-----IVYRGIS---LVILLVCAIALF-----FITKQDFYHHQKWEIRE 238
Db 238 KIIAFSTSQLGMLTIVTIGINOPYLAFLHCHTHAFKAMLFMCSGSIITHSLNDE-ODIRK 356
QY 239 KAETDRPSPSLMTFTTCYQGL-----FNWVFCFAVLT 273
Db 357 MGLLENAMP---FTTSLIIGSLATGIPFLTGFYSKDLIETANTSTNAW---ALLMT 410
QY 274 IPVVMTV--TTR-----GDSGF--LNKIMSENDEIYTLTSLFVFNLFAAIGSIVAS 322
Db 411 LIATSLTAVYSTRIFFALLGQPRELPLTSINENNPFINSIKRLIGSIFA--GFFISN 468
QY 323 KHWPT-----PRYLFAILRALFPFPFFC-----NRYVQTRAYPVFFEST 365
Db 469 NI-YPTTVPEMTPTYMK-----LTALAVTILGFTLALSLMTHNLKLE-----HST 515
QY 366 DIFVIGGIAMSFHGYLSALAMGYTPNVVPSHYSRFAAQLSVCTLMVGLLTGLW----- 420
Db 516 NVF-----KFSN-----LGIYPTTM--HRLPPLANLSMSQKSASLLDLSIWLENIL 560
QY 421 PVVIEHFVDKPSIL 434
Db 561 PKSISQFQMKTSIL 574

RESULT 14
TCR2_BACSU STANDARD; PRT; 459 AA.
AC P14512;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Tetracycline resistance protein.
GN TET.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RA Noruchi N., Aoki T., Sasatsu M., Kono M., Shishido K., Ando T.;
RT "Determination of the complete nucleotide sequence of pNS1, a
RT staphylococcal tetracycline-resistance plasmid propagated in Bacillus
RT subtilis."
RL FEMS Microbiol. Lett. 37:283-288(1986).
CC -!- FUNCTION: RESISTANCE TO TETRACYCLINE BY AN ACTIVE TETRACYCLINE
CC EFFLUX. THIS IS AN ENERGY-DEPENDENT PROCESS THAT DECREASES THE
CC ACCUMULATION OF THE ANTIBIOTIC IN WHOLE CELLS. THIS PROTEIN
CC FUNCTIONS AS A METAL-TETRACYCLINE/H+ ANTI-PORTER.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN
CC AS THE DRUG RESISTANCE TRANSLOCASE FAMILY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M16217; AAA19179.1; -
CC PIR; S42238; S42238.
CC InterPro; IPR001411; TCR_TetB.
CC InterPro; IPR003662; sub_transporter.
CC Pfam; PF00083; sugar_ttr; 1.
CC PRINTS; PR01036; TCR_TETB.
```

```
KW Antibiotic resistance; Transmembrane; Transport; Symport; Plasmid.
FT TRANSMEM 12 33 POTENTIAL.
FT TRANSMEM 81 100 POTENTIAL.
FT TRANSMEM 111 129 POTENTIAL.
FT TRANSMEM 140 162 POTENTIAL.
FT TRANSMEM 165 185 POTENTIAL.
FT TRANSMEM 201 221 POTENTIAL.
FT TRANSMEM 223 240 POTENTIAL.
FT TRANSMEM 256 276 POTENTIAL.
FT TRANSMEM 297 317 POTENTIAL.
FT TRANSMEM 324 344 POTENTIAL.
FT TRANSMEM 346 365 POTENTIAL.
FT TRANSMEM 432 451 POTENTIAL.
SQ SEQUENCE 459 AA; 50695 MW; FE53C55535372B41 CRC64;

Query Match 4.9%; Score 113; DB 1; Length 459;
Best Local Similarity 18.1%; Pred. No. 0.59;
Matches 82; Conservative 79; Mismatches 133; Indels 160; Gaps 20;

QY 22 YNIVYVLVILGVGLLPWNFEITIAPEYVYVWFKPDGVETWYKSEFMSLTIGSOLPN 81
Db 13 YSVLFWLCLLSFSSVLN--EMVYNVSLPDIAHFNTTTCITWVNTAYMLTFSIGTAVYG 70
QY 82 ASINVFNL-FLIAGPLIYRVFAPVCFNIVNLTIILIVILEPTSDMSWFFWTLGMA 140
Db 71 KLSDDYINIKKLLIIG-----ISLCLGLLIAFI-----GHNHFFILIFG-- 109
QY 141 TSINFSNGLYENSVYGVG-GDFPHT-----YIGALL-----IGNNI 175
Db 110 -----RLVQGVGSAAPPSLMVYVARNITRKKGKAFGTGISTVALGEGLGPSI 158
QY 176 CGL-----LITVVKIGVTVFLENDEPLVAIVYFGISL-----VILLVC 213
Db 159 GGIHAIYHWSYLLILPMTIVTI-----PELIKVMVPKSKNTKNTDIVIGVILMSI 209
QY 214 AIALFFITKQDFYHHQKGMREKAEPRSPSILMTFTTNCYQGLFNVMVFCFAVLT 273
Db 210 SIICFML-----FTTNVNTWTFLLFTI----- 231
QY 274 IPVVMTVTTRGDSGLNKMSEN-DEIYTLTSLFVFNLFAAIGSIVASKIHWPTPYL 332
Db 232 FFVIFIKHISRVSNPFPINPKLGKIPFMLGSLFSGGLIFSIVAGFSMWVPMY--KTIYHV 289
QY 333 KPAIILRALFIP-----FFFCNVYRVQTRAYPVFFESTDIFVIGGIAMSFHGYLSAL 385
Db 290 NVATTGNSVIFPGTMSVIVFGYGGFLVDRKG-----SLFVILGSLSIS-----ISFL 338
QY 386 AMGYTPNVVPSHYSRFAAQLSVCTLMVGLLTGGL 419
Db 339 TIAF-----FVEFSMWLT--TFMFIVMGGL 362

RESULT 15
YDNK_LACLC STANDARD; PRT; 614 AA.
AC P42377;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 70.0 kDa protein in dnaK 3'region (ORF4).
OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1359;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MG1363;
RX MEDLINE=94172317; PubMed=8126443;
RA Eaton T.J., Shearman C.A., Gasson M.J.;
RT "Cloning and sequence analysis of the dnaK gene region of Lactococcus
RT lactis subsp. lactis."
RL J. Gen. Microbiol. 139:3253-3263(1993).
```

```
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X76642; CAA54090.1; -.
KW Hypothetical protein.
SQ SEQUENCE 614 AA; 69869 MW; B1146136D09B89D9 CRC64;

Query Match          4.9%; Score 113; DB 1; Length 614;
Best Local Similarity 19.3%; Pred. No. 0.78; Indels 196; Gaps 26;
Matches 98; Conservative 66; Mismatches 149;

QY 21 KYNIVYVLVILVGFGLLPWNMFITIA---PEYVYNY-----54
DB 16 KYNFEFYIIL-----FITLATFPYRCNFHAGNDFAFNARVMSTISALKDG 64

QY 55 -----WFKPDGVB-----TWYSEFGMSLIGSOLPNASINVFNLIIAGPLIYRVFAPVC 106
DB 65 QVIOQDPNALSGFGYAW--NEFYG-----PLPTYFISVIK-FIVKWSLSFSLFYSLC 115

QY 107 FNVNLTILILIVLEPTDSMSFFWVTLCMATSFNSNGLYENSYGVGGDPHTYI 166
DB 116 LFISGIFTFNSSFULKDHNTS-KLFGLLAVALFT---FSNSTYINLYYYANPSQLALL 171

QY 167 GALLI--GNN-----ICG-----LLITVVVKIGVT 188
DB 172 FVILLFWGMNKNKRSFAFLVAFVAGAGLPLSHVTITICTLPFVLLYLLLIKKG-- 229

QY 189 YFLNDEPKLVAIVYRGISLVILLVCAIALFF-----ITKODFYH-----227
DB 230 -NLKENIKIIGIGFLSVTSAI---GLSAFFLPLENLKSGIYNVNSDFSRSGWNNI 284

QY 228 -YHHOKGMEIREKAETDRSPSILWTF-----TNC-YGOLFNVWFCFAYT 271
DB 285 AYFOGKEPLYKIEFYKFPSSLFVFLVFIFISLINFKNKTNAKYSLIFS--CFSLV 341

QY 272 LTIFPVMVTVTRGDSGFLNKIMSENDEIYTLTS---FLVFNLFAAIG-----SIVAS 322
DB 342 LVLMQLPI-----FPWKLF-----IFTIVQDPARFSTLFGLSALSLVLILPILLD 388

QY 323 KIHWPTRPYLFAILLRALFIPFFFCNRYVQTRAYPVF-----FESTDIF 368
DB 389 KISGKTSYLYLTIGLVIFSLGFAEFNR-RIOKGSQPLFASQAQSLLNKTPFNMYMENPDSI 447

QY 369 VIGGIAMSFSGYLSALAMGYTPNVVPSH 397
DB 448 AIG-----EYLPQVIGSH 460
```

Search completed: July 23, 2002, 16:43:23  
Job time: 496 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 15:41:32 ; Search time 44.23 Seconds  
(without alignments)  
942.862 Million cell updates/sec

Title: US-08-816-011f-63  
Perfect score: 2294  
Sequence: 1 MVIINRSNTYAVEQAFPRDKYNIYVWLVGVGLLPWNMFIITIAPEYYVNYWFKPDG 434  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Keeping first 45 summaries

Database :

PIR\_71: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	2265	98.7	434	2 T16080	hypothetical prote
2	716.5	31.2	461	2 T28006	hypothetical prote
3	706.5	30.8	450	2 T23528	hypothetical prote
4	574.5	25.0	800	2 T23247	hypothetical prote
5	313.5	13.7	428	2 T01489	hypothetical prote
6	249.5	10.9	408	2 C86156	T14P4.9 protein -
7	214	9.3	326	2 JC4196	36k hydrophobic nu
8	209.5	9.1	327	2 JC4195	36k hydrophobic nu
9	196.5	8.6	513	2 T21887	hypothetical prote
10	182.5	8.0	517	2 S36712	FUN26 protein - ye
11	180	7.8	143	2 T22164	hypothetical prote
12	144	6.3	418	2 D85064	hypothetical prote
13	138.5	6.0	418	2 E85064	hypothetical prote
14	134	5.8	432	2 H64383	Na+ transporter -
15	130	5.7	425	2 E97095	MDR-type permease
16	128.5	5.6	382	2 G96641	hypothetical prote
17	127.5	5.6	581	2 G90539	hypothetical prote
18	126	5.5	418	2 F83986	transporter BH2694
19	125.5	5.5	674	2 T21217	hypothetical prote
20	121	5.3	492	2 B90373	sugar transport re
21	120	5.2	330	2 C95844	probable sugar ABC
22	118	5.1	498	2 F89861	Na+/H+-antiporter
23	117.5	5.1	506	1 D64048	iron (III) ABC tra
24	117.5	5.1	559	2 S62503	inorganic phosphat
25	117	5.1	463	2 F90285	metabolite transpo
26	115.5	5.0	413	2 H95041	polysaccharide tra
27	115	5.0	456	2 A90013	conserved hypoteht
28	115	5.0	461	2 B83601	probable transport
29	114.5	5.0	449	2 C75053	DNA damage-inducib

ALIGNMENTS

RESULT 1

T16080  
hypothetical protein F16H11.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Jun-2000  
C:Accession: T16080  
Submitted to the EMBL Data Library, April 1996  
Accession: The sequence of C. elegans cosmid F16H11.  
A:Reference number: Z18458  
A:Accession: T16080  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-434 <WUX>  
A:Cross-references: EMBL:U55376; NID:gl280130; PID:gl280131; PIDN:AAA98003.1; GSPDB:G  
A:Experimental source: strain Bristol N2; clone F16H11  
C:Genetics:  
A:Gene: CESP:F16H11.3  
A:Map position: X  
A:Introns: 49/3; 94/2; 190/1; 256/3; 326/3; 392/2  
C:Superfamily: Caenorhabditis elegans hypothetical protein ZR809.4

Query Match 98.7%; Score 2265; DB 2; Length 434;  
Best Local Similarity 99.1%; Pred. No. 2e-172;  
Matches 430; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy	1	MVIINRSNTYAVEQAFPRDKYNIYVWLVGVGLLPWNMFIITIAPEYYVNYWFKPDG	60
Db	1	MVIINRSNTYAVEQAFPRDKYNIYVWLVGVGLLPWNMFIITIAPEYYVNYWFKPDG	60
Qy	61	VETWYSKFMGSLTGSOLPNASINVFNLFLIAGPLIYRVFAPVCFNVLNLTILVLI	120
Db	61	VETWYSKFMGSLTGSOLPNASINVFNLFLIAGPLIYRVFAPVCFNVLNLTILVLI	120
Qy	121	VLEPTEDSMNSFFWVTLGMATSIENFNGLYSVGVGDPTTYIGALLIGNNICGLLI	180
Db	121	VLEPTEDSMNSFFWVTLGMATSIENFNGLYSVGVGDPTTYIGALLIGNNICGLLI	180
Qy	181	TVVKIGVTVFLNDEPKLVAIVFGISLVLLVCAITALEFFITKQDFYHHQKMEIREKA	240
Db	181	TVVKIGVTVFLNDEPKLVAIVFGISLVLLVCAITALEFFITKQDFYHHQKMEIREKA	240
Qy	241	ETDRSPSILWTTFTNFCYQQLFNWFCFAVLTITFPVMMVTTRGDSGFLNKIMSENDEI	300
Db	241	ETDRSPSILWTTFTNFCYQQLFNWFCFAVLTITFPVMMVTTRGDSGFLNKIMSENDEI	300
Qy	301	YTLTSTFLVFNLFRAIGSVASKIHWPTPRYLKFAILLRALFIPFEFCNRYVQTRAYPV	360
Db	301	YTLTSTFLVFNLFRAIGSVASKIHWPTPRYLKFAILLRALFIPFEFCNRYVQTRAYPV	360
Qy	361	FFESTDIEVIGGIAMSFSGVLSALAMGYTPNVVPSYRFAAQLSVCTLMVGLLTGGIWL	420





Db 273 FP-----GFAENLKSQLLQSWYPILLITVYNISDFVGKSLFALYLWQSIKAT 321  
QY 334 FAIIILRALFIPEFFPCNRYQVTRAYPVFFESDIFVIGIAMSFSGHYLSALAMGYTPNV 393  
Db 322 WACIVRLLFYPLFSAC-----LRGPKMLRTVPVVLTFMGLTNGYLTSLVLMIMAPKT 375  
QY 394 VPSHYSRFAAQLSVCTLMVGLLTGG-----LW 420  
Db 376 VHASEAELAIFFMVVFLGLVCGSGVIGLW 406  
  
RESULT 7  
JC4195  
36k hydrophobic nucleolar protein - human  
C:Species: Homo sapiens (man)  
C:Date: 10-Sep-1995 #sequence\_revision 27-Oct-1995 #text\_change 21-Jul-2000  
R:Accession: JC4195  
R:Williams, J.B.; Lanahan, A.A.  
A:Title: A mammalian delayed-early response gene encodes HNP36, a novel, conserved nucleolar protein  
A:Reference number: JC4195; MUID:95367016  
A:Accession: JC4195  
A:Molecule type: mRNA  
A:Residues: 1-326 <WIL>  
A:Cross-references: EMBL:X86681; NID:951266; PIDN:CAA60380.1; PID:951267  
A:Experimental source: heart  
C:Comment: This protein has a role in the growth response and participates in the proliferation  
C:Genetics:  
A:Gene: der12  
A:Start codon: AUG  
C:Keywords: nucleoprotein; phosphoprotein  
F:54/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #status predicted  
F:107/Binding site: phosphate (Thr) (covalent) (by casein kinase II) #status predicted

Query Match 9.3%; Score 214; DB 2; Length 326;  
Best Local Similarity 22.0%; Pred. No. 1.2e-09;  
Matches 71; Conservative 52; Mismatches 139; Indels 60; Gaps 8;  
  
QY 143 INFNSGLYENSVYGVGGDPPTTYIGALLIGNICGLLITVVKI-----GVTVFLNDEPKL 197  
Db 7 INSFSAVLQGSFUGQLTMPSTYSLFSGQGLAGIFAALAMLLSMASGV-----DAET 60  
  
QY 198 VAIVYF---GISVLVLLCAIALFFITKQDFY---HYHQKGMREIRA----- 240  
Db 61 SALGYFITPYVGLMSIVCYLSPLHLKARYLLANKSSQAQAELETKAELQSDENGIP 120  
QY 241 -----ETDRSPSILWTTFTNCYQQLFNWVFCFAVTLTIPFV 277  
Db 121 SSPQKVALTLDLLEKEPESEPEPKPKPSVFTVFGKIMLTALCLVLVFTVTLVSFPA 180  
QY 278 MMTVTTRGDSGFLNKIMSENDEIYTLTSLFVNLFAAIGSIVASKIHP--TPRYLKFA 335  
Db 181 ITAMVTSST-----PKWQOFFNPICCFLLFNIMDMWLGRLSLTYFLWPDSDRLPL 234  
QY 336 IILRALFIPEFFPCNRYQVTRAYPVFFESDIFVIGIAMSFSGHYLSALAMGYTPNVVP 395  
Db 235 VCLRFLFVPLFVLMCHVPORSR-LPILFQDAYFITFMLFAVSNGLYLSLTMCLAPROVL 293  
QY 396 SHYSRFAAQLSVCTLMVGLLTG 417  
Db 294 PHEREVAGALMTFFLALGLSCG 315  
  
RESULT 8  
JC4195  
36k hydrophobic nucleolar protein - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 03-Sep-1995 #sequence\_revision 27-Oct-1995 #text\_change 05-Nov-1999  
R:Accession: JC4195  
R:Williams, J.B.; Lanahan, A.A.  
A:Title: A mammalian delayed-early response gene encodes HNP36, a novel, conserved nucleolar protein

A:Reference number: JC4195; MUID:95367016  
A:Accession: JC4195  
A:Molecule type: mRNA  
A:Residues: 1-327 <WIL>  
A:Cross-references: EMBL:X86682; NID:951302; PIDN:CAA60381.1; PID:951303  
C:Comment: This protein has a role in the growth response and participates in the proliferation  
C:Genetics:  
A:Gene: der12  
A:Start codon: AUG  
C:Keywords: nucleoprotein; phosphoprotein  
F:82-159/Domain: hydrophilic #status predicted <HYD>  
F:54/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #status predicted  
F:107/Binding site: phosphate (Thr) (covalent) (by casein kinase II) #status predicted

Query Match 9.1%; Score 209.5; DB 2; Length 327;  
Best Local Similarity 22.4%; Pred. No. 2.8e-09;  
Matches 76; Conservative 49; Mismatches 134; Indels 81; Gaps 12;  
  
QY 131 WPFWVTLGMAISINFSNGLYENSVYGVGGDPPTTYIGALLIGNICGLLITVVKI----- 185  
Db 5 WF-----INSFCAVLQGSFUGQLTMPSTYSLFSGQGLAGIFAALAMLSLAS 54  
QY 186 GVTYFLNDEPKLVAIVYF---GISVLVLLCAIAL-----FFI-----TK 222  
Db 55 GV-----DAQTSALGYFITPCVIGILLSIVCYLSPLHLKARYLYLTKLSQAQTOLETK 108  
QY 223 QDFYHYHQKGMREIR-----EKAETDRSPSILWTTFTNCYQQLFNWV 266  
Db 109 AELLQADEKNGVPISPQQAASPTLDLDPEKEPEPEPKPKPSVFTVFGKIMLTALCLVL 168  
QY 267 CFAVTLTIPFV--MTVTTRGDSG---FLNKIMSENDEIYTLTSLFVNLFAAIGSIV 320  
Db 169 VFTVTLVSFPAITAMVTTSSNSPGKWLFFNP-----CCFLFNWMDWLGRLSL 217  
QY 321 ASKIHP---TPRYLKFAIILRALFIPEFFPCNRYQVTRAYPVFFESTDIFVIGIAMSF 377  
Db 218 TSYFLWPDSDSOLLPLLVCLRFLEVPFLMCHVPQHAR-LPILFQDAYFITFMLFAV 276  
QY 378 SHGYLSALAMGYTPNVSPSHYSRFAAQLSVCTLMVGLLTG 417  
Db 277 SNGYLVSLTMCLAPROVLPHEREVAGALMTFFLALGLSCG 316  
  
RESULT 9  
T21887  
hypochemical protein F36H2.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T21887  
R:Steward, C.  
submitted to the EMBL Data Library, October 1996  
A:Reference number: Z19483  
A:Accession: T21887  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-513 <WIL>  
A:Cross-references: EMBL:Z81078; PIDN:CAB03075.1; GSPDB:GNO0019; CESP:F36H2.2  
A:Experimental source: clone F36H2  
C:Genetics:  
A:Gene: CESP:F36H2.2  
A:Map position: 1  
A:Introns: 28/3; 56/1; 93/2; 163/3; 277/3; 314/1; 341/3; 384/2; 407/2; 458/2; 480/3  
  
Query Match 8.6%; Score 196.5; DB 2; Length 513;  
Best Local Similarity 21.4%; Pred. No. 5e-08;  
Matches 95; Conservative 77; Mismatches 159; Indels 113; Gaps 21;  
  
QY 31 LVGFGVLLPWNMFITIAPEYYVNYWFKPDGV-----ETWYSKPEMGSLITGSQLP 80  
Db 11 IVSISMFNFNAFMN-AHEY---FYIKLRNVTNDEDPVNTTDTWFIKR-----QIRDDVP 61

Db	172	-PKWFNFIMLVLVVISMGMTATQNGIMAIANVFGSEYSQGVWVGOAVGLPSLVLF	230
Qy	187	VTFLENDEPKLVA--IVYGISVILLVCAIALFFITK-----	222
Db	231	LAFIENSVSVTGGILLYFFFTTLVVTIC--VVMFSVSKISRKVNENNVNVEDGHITDVLIG	289
Qy	223	-----QDFYHYH-----HOKGMEIREKAETDRPSPSILWTFPTNC	257
Db	290	SLRSNEEIRIVGRIDQMEDHRRRTGTRDDNDECEELQKVPFE-----VLF	341
Qy	258	YGQLENVWFCEAVTLTTPVVMVTVTRGDSGLNKINSENDEIYTLTSLFVNENLFAAIG	317
Db	342	KYLSLSTFTFVVTL--VFPVFASAT-----YVTGPLSNAQYIPLI--FTLNLGDLYG	392
Qy	318	SIVASKIHWP-----TPRYLKFAIILRALFPFF	346
Db	393	RVIAAD--WPMFRDQKFTPRKFTIYSLLRVAAIPLF	425

```

RESULT 11
T22164
hypothetical protein F44D12.9 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T22164
P:Coles, L.

```

A;Reference number: Z19523  
A;Accession: T22164  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-143 <WTL>  
A;Cross-references: EMBL:Z68298; PID:CAA92605.1; GSPDB:GN00022; CESP:F44D12  
A;Experimental source: Clone F44D12  
C;Genetics:  
A;Gene: CESP:F44D12.9  
A;Map position: 4  
A;Introns: 19/1; 37/2

```

Query Match      7.88;   Score 180;   DB 2;   Length 143;
Best Local Similarity 40.08;   Pred. No. 2.5e-07;
Matches 32;   Conservative 18;   Mismatches 26;   Indels 4;   Gaps

Qy 14 QEAPPROKYNIVYVLVILGFGVLLPNWMTITTAPEYVNVYWFKPDGVETWYSKEFMGSL 73
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 67 KESSETRDWILVYLIIFTHGCMGLMSHWMTITTAQYYHYDWFN-----NTNYQDSFMSII 122
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Qy 74 TIGSOLPNASINVENFLII 93
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 123 GVTSGIPIVNGIMILNTIVVM 142

RESULT 12
D85064
hypothetical protein AT4g05120 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)

```

C:Accession: D85064  
 R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium,  
 Nature 402, 769-777, 1999  
 A>Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana

A:Accession: D85064  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-418 <STO>  
A:CROSS-references: GB:NC\_001368; NID:g7267271; PIDN:CAB81054.1; GSPDB:GN010900001; MIM:103050  
C:Genetics:  
A:Gene: AT4q05120

```
n, map position: 4
```

Query Match 6.3%; Score 144; DB 2; Length 418;  
Best Local Similarity 20.6%; Pred. No. 0.00059;  
Matches 96; Conservative 76; Mismatches 159; Indels 134; Gaps 23;

QY 18 PRKYNIYVWVI---LVGQVLLPWNMFITIAPEYV-VNYWFKPDGVETWYSKEF-MGS 72  
DB 9 PPELQKQYQAMVVCILGIGSLVSNMMLTIADYYVQVFPDYHPSRVLTFLVYQPEALGT 68  
QY 73 LTI-----GSOLPNASINVENFLIIAGPLIYRVFAPVCNIVNLTILILVIVLEPTEDS 128  
DB 69 ILILAYHESKINTRKRNLIYIL-----FTISTELLIVLD----- 103  
QY 129 MSFEFWTLGMAT-----SINFNGLYENSUVG--VGG-----DEPHTYIG 167  
DB 104 -----LATKGGGIGPYIGLCVAVNASVGLADATVQGGMIGDLSLPCPELVOSFMG 153  
QY 168 ALLIGNNICGLLTVTKVIGTYFLNDEPKLVAIVYFGISVLIVLCAIALFFI-TKQDFY 226  
DB 154 GLAVSGALTSALRLITKAAFEK-TNDGPRKGAAMFLAISTCIELLCVELYAVVFPKLPV 212  
QY 227 HYHQKQMEIREK-----AETDRPSPSILWTTFTNCYQGLFNW 265  
DB 213 KYRRKAASEGSKTVSADLAAAGIQNSDLTDDDSKNQRLSNKELLIQNIDYAVNLFYI 272  
QY 266 FCFPAVLTITFPVNMVTTTRGDSGFLNKINSEN--DEIYTLTSLFVNLFAAIGSIVASK 323  
DB 273 VC---TLSIFP-----GFLYENTGQHLGDWYALVL-VAMYNCDLVGR----- 312  
QY 324 IHWPTRYLK-----FAILRLALFIPFFFCNRYVOTRAYPVFFESTDIFVIGGIAM 375  
DB 313 -YTPLVKWLKIENRKLITIAVLSRYLLIPAFYF-TAKYGDQGMIML-----ISVLG--- 362  
QY 376 SPFHGYLSALAMGYTNNVPVSHYSRFAAQLSVCTLMVGLLTGGLW 420  
DB 363 -LTNGHLTVCM---TIAPKGY-KGPEQNALGNLLVFLGLGIF 401

RESULT 13  
E85064  
hypoetical protein AT4g05130 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cross)  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 16-Feb-2001  
C:Accession: E85064  
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring Nature 402, 769-777, 1999  
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
A:Reference number: A85001; MUID:20083488  
A:Accession: E85064  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-418 <STO>  
A:Cross-references: GB:NC\_001268; NID:g7267272; PIDN:CAB81055.1; GSPDB:GN00140  
C:Genetics:  
A:Gene: AT4g05130  
A:Map position: 4

Query Match 6.0%; Score 138.5; DB 2; Length 418;  
Best Local Similarity 21.0%; Pred. No. 0.0016;  
Matches 95; Conservative 73; Mismatches 174; Indels 111; Gaps 23;

QY 21 KYNIVVWLIVLFGVLLPWNMFITIAPEYV-VNYWFKPDGVETW-YSKFEFGSLRIGSQ 78  
DB 16 KQAMVVCIL-GIGSLFSWNSMLTIADYYVQVFPDYHPSRVFTLYIQPIALGTIMI--- 71  
QY 79 LPNASINVENFLIIAGPLIYRVFAPVCNIVNLTILILVIVLEPTEDSMSFEWTLG 138  
DB 72 LAYRESKISTRRLTGTYLF-----TISTELLIVLDLT-----TKG 108  
QY 139 -----MATSNFNGLYENSIVY-GVGDRP-----HTVIGALLIGNNICGLLITVVKG 186  
DB 109 HGGIGHYIVLCTVASFGLADATVKGLVGDLSLPCPELIQSYMGSMAGALTSVLR- 167

QY 187 VTYFL---NDEPKLVAIVVFGISVLIVLCAIALFFI-TKQDFYHYHOKG----- 233  
DB 168 ITRAAAEKSNNSURKAMFLAISTFIELLCVLIVAYVFPKLVIVKYRKAASEGSKTV 227  
QY 234 -----MEIREKAETD-----RPSPSILWTTFTNCYQGLFNWFCFAVTLTIFFVM 278  
DB 228 VADLAAAGIQNSLSDSDSKNQMLRKKELLQN-----IDHAVNLFILVVLTLISIFF-- 280  
QY 279 MTVTTRGDSGFLNKINSEN--DEIYTL--TSLVNLNFAAIGSIVASKIHW---PTPRY 331  
DB 281 -----GFLYENTGQHLGDWYALIVATYFNWDLFGRYAPLV---KWLKLENRKA 327  
QY 332 LKFAIILRALFIPFFFCNRYVOTRAYPVFFESTDIFVIGGIAMSPSHGYSLSALAMGYTP 391  
DB 328 LTATVTRVFLVPAFYF-TAKYGDKGMMIMLV-----JLGLTTGHLTVCIIMTIAP 377  
QY 392 NVVPVSHYSRFAAQLSVCTLM---VGLLTGGLW 420  
DB 378 NGYKGPKEKNALGNLLVVFIILGGAUVGSLGWLW 410

RESULT 14  
H64383  
Nat transporter - Methanococcus jannaschii  
C:Species: Methanococcus jannaschii  
C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 31-Dec-2000  
C:Accession: H64383  
R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak ; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.  
Science 273, 1058-1073, 1996  
A:Authors: kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc A:Reference number: A64300; MUID:96337999  
A:Accession: H64383  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-432 <BUL>  
A:Cross-references: GB:U67514; GB:L77117; NID:g2826304; PIDN:AAB98666.1; PID:g1591386  
C:Genetics:  
A:Map position: REV598993-597695  
C:Superfamily: probable transporter MJ0672

Query Match 5.8%; Score 134; DB 2; Length 432;  
Best Local Similarity 19.1%; Pred. No. 0.0036;  
Matches 82; Conservative 80; Mismatches 133; Indels 134; Gaps 22;

QY 66 SKFEMG-----SLTIGSOLPNASINVENFLIIAGPLIYRVFAPVCNIVNLTILIV 119  
DB 4 SKFELGIIITASLIFGSSLPDIYKGI--VILLIVACLWFFELLPLPVTSLAIPIMAVFL 61  
QY 120 IVLEPTEDSMSNF-----FWWTLG---MATSNFNSN--GLYENSIVYGVGGDFPHF----- 164  
DB 62 GIFN-LKEALTVAHPHPIIFLFGGFMALQAKNHNLDKFIAYKLLNYGDKFTCFMLFL 120  
QY 165 --YIGALLIGNNICGLLITVVKIGVTV-----FLNDEPKL 197  
DB 121 SAYFLSMWISNTSATLILPILALGLLHKTKGLRDLFLLGVAYSASIGGIATIGSGPNA 180  
QY 198 VAIVY-----FGISVLIVLCAIALF-----FITKQDFYHYHQKQMEIREKA 240  
DB 181 IASSYLDYGFSSWFKVGFPPISLLFLICLTLYIYFKWIPKED-----IAIQARM 231  
QY 241 ETRDRPSPI-----LWTTFTNCYQGLFNWFCFAVTLTIFFVMVMTTRGDSGFLN 291  
DB 232 ELSRNAYKLLVIFVLIASLW-IISDYLSEIFNVQY-FDSVIAIFAILL-----FVF 281  
QY 292 KIMSEND---EITYTL-----TSFLVNLFAAIGSIVASKIHWPTPR 330  
DB 282 NLVEVNDFKKIDWGTILFGLGALCLGVIVKSGANTFLSEKLIAILGNL-----TPI 333



22 1EVO13QUG1YRUL3K31E0RLEVFANNAANEFFGFINIFGRFSALG - FFMVGLVS 368

Search completed: July 23, 2002, 16:36:11  
Job time: 3279 sec

**THIS PAGE BLANK (USPTO)**

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	343.5	15.0	456	4	US-09-058-389A-4	Sequence 4, Appli
2	326.5	14.2	456	4	US-09-058-389A-2	Sequence 2, Appli
3	214	9.3	326	4	US-09-058-389A-3	Sequence 3, Appli
4	181	7.9	247	4	US-09-058-389A-8	Sequence 8, Appli
5	111.5	4.9	398	1	US-08-097-938-4	Sequence 4, Appli
6	111.5	4.9	398	1	US-08-476-000-4	Sequence 4, Appli
7	111.5	4.9	398	1	US-08-472-840-4	Sequence 4, Appli
8	111.5	4.9	398	2	US-08-476-976-4	Sequence 4, Appli
9	111.5	4.9	398	3	US-08-474-410-4	Sequence 4, Appli
10	111.5	4.9	398	4	US-08-486-673B-4	Sequence 4, Appli
11	109.5	4.8	397	4	US-08-486-673B-63	Sequence 63, Appli
12	108.5	4.7	398	1	US-08-097-938-6	Sequence 6, Appli
13	108.5	4.7	398	1	US-08-476-000-6	Sequence 6, Appli
14	108.5	4.7	398	1	US-08-472-840-6	Sequence 6, Appli
15	108.5	4.7	398	2	US-08-476-976-6	Sequence 6, Appli
16	108.5	4.7	398	3	US-08-474-410-6	Sequence 6, Appli
17	108.5	4.7	398	4	US-08-486-673B-6	Sequence 6, Appli
18	106.5	4.6	395	1	US-08-097-938-2	Sequence 2, Appli
19	106.5	4.6	395	1	US-08-097-938-5	Sequence 5, Appli
20	106.5	4.6	395	1	US-08-476-000-2	Sequence 2, Appli
21	106.5	4.6	395	1	US-08-476-000-5	Sequence 5, Appli
22	106.5	4.6	395	1	US-08-472-840-2	Sequence 2, Appli
23	106.5	4.6	395	1	US-08-472-840-5	Sequence 5, Appli
24	106.5	4.6	395	2	US-08-476-976-2	Sequence 2, Appli
25	106.5	4.6	395	2	US-08-476-976-5	Sequence 5, Appli
26	106.5	4.6	395	3	US-08-474-410-2	Sequence 2, Appli
27	106.5	4.6	395	3	US-08-474-410-5	Sequence 5, Appli

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201-487-5800  
TELEFAX: 201-343-1684  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 456 amino acids  
TYPE: amino acid  
STRADEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
DESCRIPTION: hENT2  
HYPOTHETICAL: NO  
FRAGMENT TYPE:  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
US-09-058-389A-2

Query Match 14.2%; Score 326.5; DB 4; Length 456;  
Best Local Similarity 22.9%; Pred. No. 1.8e-24;  
Matches 106; Conservative 71; Mismatches 200; Indels 85

[illegible]

```

RESULT      3
US-93-058-389A-3
Sequence 3, Application US/09058389A
Patent No. 6130065
GENERAL INFORMATION:
APPLICANT: Belt, Judith A.
APPLICANT: Crawford, Charles R.
APPLICANT: Patel, Divyen
TITLE OF INVENTION: A NITROBENZYL
TITLE OF INVENTION: (NBMR)-INSE
TITLE OF INVENTION: PROTEIN, NUC
TITLE OF INVENTION: USE
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESS: David A. Jackson, Esq.
STREET: 411 Hackensack Ave, Cor
CITY: Hackensack

```

[illegible]

```

RESULT      2
US-09-058-389A-2
; Sequence 2, Application US/09058389A
; Patent No. 6130065
; GENERAL INFORMATION:
; APPLICANT: Belt, Judith A.
; APPLICANT: Crawford, Charles R.
; APPLICANT: Patel, Divyen
; TITLE OF INVENTION: A NITROBENZYL MERCAPTOPYRINERIBOSIDE
; TITLE OF INVENTION: (NBMPR)-INSENSITIVE, EQUILIBRATIVE, NUCLEOSIDE TRANSPORT
; TITLE OF INVENTION: PROTEIN, NUCLEIC ACIDS ENCODING THE SAME AND METHODS OF
; TITLE OF INVENTION: USE
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; STREET: Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/058,389A
; FILING DATE: April 9, 1998
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1340-1-013N

```





```

; Sequence 4, Application US/08472840
; Patent No. 5763575
; GENERAL INFORMATION:
; APPLICANT: SUNDELIN, JOHAN
; APPLICANT: SCARBOROUGH, ROBERT M.
; TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR, ITS AGONISTS
; TITLE OF INVENTION: AND ANTAGONISTS, AND NUCLEIC ACIDS ENCODING THE RECEPTOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,840
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/390,301
; FILING DATE: 25-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: ADLER, REID G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: 2803-0006.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 398 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-472-840-4

Query Match          4.9%; Score 111.5; DB 1; Length 398;
Best Local Similarity 18.5%; Pred. No. 0.0045;
Matches 75; Conservative 54; Mismatches 131; Indels 145; Gaps 18;

QY   61 VETWYS-KFMGSLTIGSOLPNASINVNLFLLIAGPLIYRVFAPVCFNIVNLTIILV 119
    ||| : || | : | : || : || : || : || : || : || : || : || : || : ||
Db   56 VETVFVSDFSAVLTKG-----LTTVFLPIVVTV-----F 87
QY   120 IVLEPTEDSMSEFF-----WVTLMGATSNFSNGLYENS 153
    || : || : || : || : || : || : || : || : || : || : || : || : ||
Db   88 VGLPNSGMALWFLFRTRKKHPAVIYMANIALADLLSVIMFPCLKIAHYHHGNMWYGEA 147
QY   154 YGVGGDFPHTYIGALLIGNCGLLITVKIGVTFYNLDEP----KLVAIVYFGISLVI 209
    || | : || | : | : || : || : || : || : || : || : || : || : ||
Db   148 LCNV-----LIG-FFYGNMYCSILFWTCLSVQRYWIWNPMGHSRKKAIAIGISLA 199
QY   210 -LLVCAIAL-FFTITKODFYHHQKGMEIREKAETDRSPSILMTFTFNCY----- 258
    || : || : || : || : || : || : || : || : || : || : || : || : ||
Db   200 WLLILLVTPLYVVKOTIF-----IPAL---NITTCHDVLPEQLLV 237
QY   259 GOLFNWFCEAVTLTIFFPVMMVTTRGDGSLF-----NKTMSNDE-----IYT 302
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db   238 GDMFNFLSLAIGVFLFPALFTA----SAYVLMIRMLRSSAMDENSEKKRAIKLIIVT 292
QY   303 LLTSLSVNFNFAAIGSIIVASKIHWPTRYLKFAILLRALFIPTFFFCFNRVOTRAYPVFE 362
    : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db   293 VLAWYL--CF-----TPSNL-----LLVVHFLIKSQCSQHVALYI 328
QY   363 ESTDIEFIGGIAMSFSGHYLSALAMGYTPNVPPSHYSRFAAQLSV 407

```

```

Qy 303 LITSELVNLPAAIGSVASKIHWPTPRYLKFAILLRALFIPFFFCFNRYRVQTFRAYPVFEF 362
      :|::| |::|| ||::|| |::|| |::|| |::|| |::|| |::|| |::||
Db 293 VLAMLYI-CP-----TPSNL-----LLVVHYFLIKSQCSHVYALYI 328
      :|::|| |::|| |::|| |::|| |::|| |::|| |::|| |::|| |::||
Qy 363 ESTDIFVIGGTAMSGHYSLSALAMGYTPNVVPVSHYSRFAOLSV 407
      :|::|| |::|| |::|| |::|| |::|| |::|| |::|| |::|| |::||
Db 329 VALCLSTLNSCIDPFVIYFVSHDFRDHAKNALLCRSRTVKOMOV 373
      :|::|| |::|| |::|| |::|| |::|| |::|| |::|| |::|| |::||

RESULT          9
US-08-474-410-4
; Sequence 4, Application US/08474410
; Patent No. 6043212
; GENERAL INFORMATION:
; APPLICANT: SUNDELIN, JOHAN
; APPLICANT: SCARBOROUGH, ROBERT M.
; TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR, ITS AGONISTS
; TITLE OF INVENTION: AND ANTAGONISTS, AND NUCLEIC ACIDS ENCODING THE RECEPTOR
; NUMBER OF SEQUENCES: 63
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,410
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/390,301
; FILING DATE: 25-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: ADLER, REID G.
; REGISTRATION NUMBER: 30,988
; REFERENCE/DOCKET NUMBER: 2803-0006.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0763
; TELEX: 90-4030
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 398 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-474-410-4

Query Match              4.9%; Score 111.5; DB 3; Length 398;
Best Local Similarity   18.5%; Pred. No. 0.0045;
Matches                  75; Conservative 54; Mismatches 131; Indels 145; Gaps 18;

Qy 61 VETWYS-KEFMGSLTIGSQLPNASINVFNFLTITAGPLIVRFAPVCFNIVNLTIILV 119
      ||||| |::|| |::|| |::|| |::|| |::|| |::|| |::|| |::||
Db 56 VETVFSDVESASVLTK-----LTVFLPIVYTV-----F 87
      ||||| |::|| |::|| |::|| |::|| |::|| |::|| |::|| |::||

Qy 120 IVLEPTEDSSWF-----WVTGLMATSFNSNGLYENS 153
      :|::|| |::|| |::|| |::|| |::|| |::|| |::|| |::|| |::||
Db 88 VVGLPNSGMALWLFERTKKKHPAVIYMANLADLLSVTFPFLKIAYHTHGNNWIYGEA 147
      :|::|| |::|| |::|| |::|| |::|| |::|| |::|| |::|| |::||

Qy 154 YVGGDPFPHTYCALLIGNNICGLLTIVVKGTYTFLENDEP----KLVAIVFGISLVI 209
      :|::|| |::|| |::|| |::|| |::|| |::|| |::|| |::|| |::||
Db 148 LCNW-----LIG-FPYGNMYSIIIFMTCLSVQRVWIVNPMGHKRKANIAIGISLA 199
      :|::|| |::|| |::|| |::|| |::|| |::|| |::|| |::|| |::||

Qy 210 -LLVCAIAL-FFTITKDQFYHHHQKMEIRKAETDRPSILWTFTTCY----- 258

```

Db	200	WLLILLVTIPLYVVKQIF-----IPAL-----NITTC	HOVLPEQLIV 237
Qy	259	GQLFNWVFCFAVTLTIPVMMVTTRGDSGL-----NKIMSENDE-----IYT 302	
Db	238	GDMFNYFLSLAIGVFLPAPFLTA-----SAYVLMIRMLRSSAMDENSEKKRRAIKLI 292	
Qy	303	LLTSFLVNLFAAIGSIVASKIHWPTPRYLKFAILRALFIPIFFFCNVRVQTRAYPVFF 362	
Db	293	VLAMYLI--CF-----TPSNL-----LLVVHYFLIKSQGQSHVYALYI 328	
Qy	363	ESTDIFVIGGIAMSFSGYLSALAMGYTPNVVPSHYSRFAAQLSV 407	
Db	329	VALCLSTLNSCIDPFVYFYVSHDFRDHAKNALLCRSVRTVKOMQV 373	
RESULT 10			
US-08-486-673B-4			
; Sequence 4, Application US/08486673B			
; Patent No. 6297026			
; GENERAL INFORMATION:			
; APPLICANT: Sundelin, Johan			
; APPLICANT: Scarborough, Robert M.			
; TITLE OF INVENTION: Nucleic Acids Encoding the C140 Receptor			
; FILE REFERENCE: 44481-5006-08-US			
; CURRENT APPLICATION NUMBER: US/08/486,673B			
; CURRENT FILING DATE: 1995-06-07			
; PRIOR APPLICATION NUMBER: US 08/097,938			
; PRIOR FILING DATE: 1993-07-26			
; PRIOR APPLICATION NUMBER: PCT/US94/08536			
; PRIOR FILING DATE: 1994-07-26			
; NUMBER OF SEQ ID NOS: 63			
; SOFTWARE: PatentIn Ver. 2.1			
; SEQ ID NO 4			
; LENGTH: 398			
; TYPE: PRT			
; ORGANISM: Homo sapiens			
us-08-486-673B-4			
Query Match 4.9%; Score 111.5; DB 4; Length 398;			
Best Local Similarity 18.5%; Pred. No. 0.0045;			
Matches 75; Conservative 54; Mismatches 131; Indels 145; Gaps			
Qy	61	VETWYS-KEFMGSLTIGSQLPNASINVFNFLIAGPLIVRVFAPVCFNIVNLITILIV 119	
Db	56	VETVFSVDSESASVLTKG-----LTTVFLPIVYIV-----F 87	
Qy	120	IVLEPTEDSSWFF-----WYTLGMATISNFSNGLYENS 153	
Db	88	VGGLPSNGMALVFLFRKTKKHPAVIYMANLALADLLSWFPLKIAYIHGNNIYGEA 147	
Qy	154	YVGVGDFPHYTCALLIGNICGLLTIVVKIGVTFLANDEP-----KLVAIVVFGISLVI 209	
Db	148	LCNV-----LIG-FYGNMYSIIIFMTCLSVORVYVWVIVNPMGHSRKKANIAIGISLAI 199	
Qy	210	-LLVCAIAL-FFITKQDFVHYHQKWEIREKAETDRSPSILWTTFTNCTY-----258	
Db	200	WLLILLVTIPLYVVKQIF-----IPAL-----NITTC	HOVLPEQLIV 237
Qy	259	GQLFNWVFCFAVTLTIPVMMVTTRGDSGL-----NKIMSENDE-----IYT 302	
Db	238	GDMFNYFLSLAIGVFLPAPFLTA-----SAYVLMIRMLRSSAMDENSEKKRRAIKLI 292	
Qy	303	LLTSFLVNLFAAIGSIVASKIHWPTPRYLKFAILRALFIPIFFFCNVRVQTRAYPVFF 362	
Db	293	VLAMYLI--CF-----TPSNL-----LLVVHYFLIKSQGQSHVYALYI 328	
Qy	363	ESTDIFVIGGIAMSFSGYLSALAMGYTPNVVPSHYSRFAAQLSV 407	
Db	329	VALCLSTLNSCIDPFVYFYVSHDFRDHAKNALLCRSVRTVKOMQV 373	

Query Match 4.9%; Score 111.5; DB 4; Length 398;  
Best Local Similarity 18.5%; Pred. NO. 0.0045;  
Matches 75; Conservative 54; Mismatches 131; Indels 145; Gaps 18;

Query Match 4.9%; Score 111.5; DB 3; Length 398;  
Best Local Similarity 18.5%; Pred. NO. 0.0045;  
Matches 75; Conservative 54; Mismatches 131; Indels 15; Gaps 18;

QY	61	VERWYS-KEPMGSLTIGSQLPNASINVFNFLIAGPLIYRVFAPCYCFNIVNLTIIILV	119
		:    :	
Db	56	VERVFSDEFSASVLTKG-----LTTVELPIVYIV-----F	87
QY	120	IVLEPTDESWEF-----WYTLGNATSNFNSGLYENS	153
		:  :  :	
Db	88	VGLPSNGMALWFLFRTKKKHPAVIYMANLADLLSVIWEPLKIAYIHGNWNIIYGEA	147
QY	154	VYVGGDPPHTYIGALLIGNICGLLITVYKIGVTYFLNDP---KLVAIYVFGISLVI	209
		:  :  :  :  :  :  :  :  :  :  :  :  :  :  :	
Db	148	LCNV-----LIG-FYGMVYCSILFTMCLSVQRIYVIVNPMGHSRKKKNTAIGISLAI	199
QY	210	-LLVCAIAL-FFITKQDFYHHQGWIREKRAETDRPSPISLWTFITNCTY-----	258



RESULT 13  
US-08-476-000-6  
; Sequence 6, Application US/08476000  
; Patent No. 5716789  
; GENERAL INFORMATION:  
; APPLICANT: SUNDELIN, JOHAN  
; APPLICANT: SCARBOROUGH, ROBERT M.  
; TITLE OF INVENTION: RECOMBINANT C140 RECEPTOR, ITS AGONISTS  
; TITLE OF INVENTION: AND ANTAGONISTS, AND NUCLEIC ACIDS ENCODING THE RECEPTOR  
; NUMBER OF SEQUENCES: 63  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 2000 Pennsylvania Ave. N.W., Ste. 5500  
; CITY: Washington







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 23, 2002, 15:36:31 ; Search time 62.22 Seconds  
(without alignments)  
774.767 Million cell updates/sec

Title: US-08-816-011f-63  
Perfect score: 2294  
Sequence: 1 MVIINRNTYAYEQEAFPRD.....LTGGLWPVVIEHFVDKPSII 434

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_032802.\*  
1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2294	100.0	434	22	AAU07630
2	2290	99.8	434	17	AA92315
3	645	28.1	458	22	AB64784
4	372.5	16.2	475	21	AA92286
5	343.5	15.0	456	19	AAW69555
6	343.5	15.0	456	21	AA92285
7	327.5	14.3	475	22	AA92285
8	327.5	14.3	475	22	AA67487
9	326.5	14.2	456	19	AAW69555
10	326.5	14.2	456	19	AAW69558
11	326.5	14.2	456	21	AA92285

12	326.5	14.2	475	21	AA92285
13	326.5	14.2	475	22	AAU07630
14	326.5	14.2	475	22	AAU07630
15	320.5	14.0	456	19	AAW69557
16	301	13.1	457	19	AAW69556
17	272.5	11.9	373	19	AAW64550
18	261.5	11.4	397	22	AA92286
19	221.5	9.7	586	22	AB64784
20	217.5	9.5	404	22	AB64784
21	214	9.3	326	21	AA92285
22	181	7.9	247	19	AAW69556
23	181	7.9	247	21	AA92285
24	180	7.8	314	20	AAW69555
25	180	7.8	314	22	AB64784
26	179.5	7.8	423	22	AB64784
27	172	7.5	248	22	AAW69556
28	125	5.4	115	22	AAW69556
29	124	5.4	521	22	AB64784
30	119.5	5.2	383	22	AB64784
31	117	5.1	866	22	AB64784
32	114.5	5.0	450	22	AB64784
33	114.5	5.0	527	21	AA92285
34	113.5	4.9	329	22	AAW69556
35	113.5	4.9	329	22	AAW69556
36	113.5	4.9	360	22	AAW69556
37	113.5	4.9	530	22	AAW69556
38	113.5	4.9	545	22	AAW69556
39	113.5	4.9	579	21	AA92285
40	113.5	4.9	582	21	AA92285
41	111.5	4.9	397	21	AA92285
42	111.5	4.9	398	16	AAW69555
43	111.5	4.9	398	17	AAW69555
44	111.5	4.9	545	21	AA92285
45	111	4.8	439	19	AAW61371

ALIGNMENTS

RESULT 1	
AAU07630	AAU07630 standard; Protein; 434 AA.
ID	AAU07630 standard; Protein; 434 AA.
XX	
AC	AAU07630;
XX	
DT	21-NOV-2001 (first entry)
XX	
DE	Caenorhabditis elegans potassium ion channel CORK protein.
XX	
KW	Transmembrane potassium ion channel protein; inward potassium flux;
KW	pest control; membrane potential; pesticide; antihelminthic; nematode;
KW	insect; CORK.
XX	
OS	Caenorhabditis elegans.
XX	
FH	Key Location/Qualifiers
FT	Modified-site /note= "N-glycosylated"
FT	Domain 26...39
FT	Modified-site /note= "Putative pore-forming H5 domain #1"
FT	Modified-site 83
FT	Modified-site /note= "N-glycosylated"
FT	Domain 146
FT	Domain /note= "N-glycosylated"
FT	Domain 150..162
FT	Domain /note= "Putative pore-forming H5 domain #2"
XX	
PN	WO200161006-A2.
XX	
PD	23-AUG-2001.
XX	
PF	14-FEB-2001; 2001WO-US04680.
XX	

```
PR 15-FEB-2000; 2000US-0503849.
XX (BADI ) BASF CORP.
XX Pausch MH;
XX WPI; 2001-536570/59.
DR N-PSDB; AAS12162.
XX
PT New polypeptide, a mutant potassium ion channel protein for improving
PT inward potassium flux under acidic conditions -
PS Example 14; Fig 9; 131pp; English.
XX
CC The invention relates to a mutant potassium ion channel protein, having
CC four membrane spanning domains and two pore forming domains, comprising a
CC mutation at the second pore forming domain. The expression of the mutant
CC protein in a cell confers improved inward potassium flux and the ability
CC to grow in the presence of potassium. Mutant proteins and their
CC corresponding polynucleotide sequences can therefore be used to improve
CC inward potassium flux into cells under acidic conditions by modulating
CC the membrane potential using therapeutic agents. The sequences may be
CC used to develop agonists and antagonists of potassium channel proteins in
CC order to control pests such as nematodes and insects. This sequence
CC represents a Caenorhabditis elegans transmembrane potassium ion channel
CC protein, CORK.
XX
SQ Sequence 434 AA;
```

```
Query Match 100.0%; Score 2294; DB 22; Length 434;
Best Local Similarity 100.0%; Pred. No. 8.7e-228;
Matches 434; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MVIINRSNTYAVEQAPRDKYNIYVWLVILVGFVLLPWNMFITIAPEYYVNYWFKPDG 60
Db 1 mviinrsntyaveqaprdkyniyvwilvlgfvgllpwnmfitiapeyyvnywfkpgd 60
Qy 61 VETWYSKEFMGSLTIGSOLPNASINVFNLFLIAGPLIYRVFAPVCFNVLNLTILILVI 120
Db 61 vetwyskefmgsltigsqlpnasinvnfllfiagpllyrvfapvcfnvltlililvi 120
Qy 121 VLEPTEDSMSPFWTGLMATSFNSNGLYENSIVYGVGDEPHTYIGALLIGNNICGLLI 180
Db 121 vleptedsmsfwvtlgmtsinfnglyensvyvgvgdftbtyigallignnicglli 180
Qy 181 TVWKIGVYFLNDEPKLVAIVYFGLSVLLVLCATALFFITKQDPRYHHQKMEIREKA 240
Db 181 tvwkigvtyflndepklvaivyfglsllvllvcatalffitkqdfyhhqkmeireka 240
Qy 241 ETDPRSPSLWTFTNCGYGFVWFCFAVTLTIPVMMVTTRGDSGFLNKMSENDEI 300
Db 241 edrpspsllwtftncyqgfwnwfcfavtltipvmmtvttrgdsfglnkmsendei 300
Qy 301 YTLTSLFVNLFAAIGSIVASKIHWPTRPYLKFALILRALFIPFFFCNRYVQTRAYPV 360
Db 301 ytltsflvnlfaaigsiavaskihwptpylklfaiilralfipfffcnyrvqtraypv 360
Qy 361 PFESTDIFVIGGIANSFSGHLSALAMGYTNVPVSHYSRAAQLSVCTLMVGLLTGGLW 420
Db 361 pfestdifvigiannsfshgylsalamgytnvpvshysrfaaqlsvctlmvlltggllw 420
Qy 421 PVVIEHFVDKPSIL 434
Db 421 pvviehfvdkpsil 434
```

```
RESULT 2
AAR92315
ID AAR92315 standard; Protein; 434 AA.
XX
AC AAR92315;
XX
```

```
DF 16-OCT-1996 (first entry)
XX
DE CORK potassium channel protein.
XX
KW CORK; potassium channel; nematode; pore-forming domain;
KW transmembrane helix; N-glycosylation site; potassium-agonist;
KW potassium-antagonist; drug screening; nematocide; anthelmintic;
KW cardiac disorder.
XX
OS Caenorhabditis elegans.
XX
FH Key Location/Qualifiers
FT Modified-site 5..7 /note= "N-glycosylation site"
FT Domain 26..39 /note= "Pore-forming H5 domain"
FT Misc-difference 33..35 /note= "G-Y/F-G motif"
FT Modified-site 81..83 /note= "N-glycosylation site"
FT Modified-site 144..146 /note= "N-glycosylation site"
FT Domain 150..162 /note= "Pore-forming H5 domain"
FT Misc-difference 156..158 /note= "G-Y/F-G motif"
XX
PN WO9613520-A1.
XX
PD 09-MAY-1996.
XX
PF 25-OCT-1995; 95WO-US14364.
XX
PR 31-OCT-1994; 94US-0332312.
XX
PA (AMCY ) AMERICAN CYANAMID CO.
XX
PI Pausch MH, Price LA;
XX
DR WPI; 1996-239450/24.
DR N-PSDB; AAT18168.
XX
PT Potassium channel genes from Drosophila melanogaster and
PT Caenorhabditis elegans - useful in assaying substances to determine
PT effects on cell growth, and in inhibiting nematode and insect pests
XX
PS Claim 13; Fig 9; 79pp; English.
XX
CC This potassium channel sequence is encoded by the CORK gene from
CC Caenorhabditis elegans, and has 2 pore-forming domains situated
CC between hydrophobic transmembrane helix domains. The CORK sequence
CC contains structural features resembling pore-forming H5 domains
CC found in potassium channels. 2 Putative pore-forming H5 domains
CC contain the G-Y/F-G tripeptide motif required for potassium
CC selectivity. The protein contains 3 asparagine-linked glycosylation
CC sites. The protein may be expressed in a heterologous host cell to
CC assay substances to determine effects on cell growth. Potassium-
CC agonists or potassium-antagonists identified by this method may be
CC used as nematocides, anthelmintics or in therapy of cardiac
CC disorders, etc.
XX
SQ Sequence 434 AA;
```

```
Query Match 99.8%; Score 2290; DB 17; Length 434;
Best Local Similarity 99.8%; Pred. No. 2.2e-227;
Matches 433; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 MVIINRSNTYAVEQAPRDKYNIYVWLVILVGFVLLPWNMFITIAPEYYVNYWFKPDG 60
Db 1 mviinrsntyaveqaprdkyniyvwilvlgfvgllpwnmfitiapeyyvnywfkpgd 60
Qy 61 VETWYSKEFMGSLTIGSOLPNASINVFNLFLIAGPLIYRVFAPVCFNVLNLTILILVI 120
```

Db	61	vwysketngstltsqldpnasinfnfliaagllyrvfapvcfnlnvntllilvli	120
QY	121	vleptedsmswffwvltlgmatsinfngslyensvtyvgvgdpphtyigallignnicgllli	180
Db	121	vleptedsmswffwvltlgmatsinfngslyensvtyvgvgdpphtyigallignnicgllli	180
QY	181	tvvkigvtyfndeplkvalvfygislvtlllvcatalffittkqdfyhyhhqkcmreka	240
Db	181	tvvkigvtyfndeplkvalvfygislvtlllvcatalffittkqdfyhyhhqkcmreka	240
QY	241	etdrpspsilwtftncygoelfnvwfcfavtltifpvmmtvttrgdsgflnkimsendei	300
Db	241	etdrpspsilwtftncygoelfnvwfcfavtltifpvmmtvttrgdsgflnkimsendei	300
QY	301	ytlltstflvfnlfaaigsvaskihwptprylkfaiillralflpffffcnrvqtravp	360
Db	301	ytlltstflvfnlfaaigsvaskihwptprylkfaiillralflpffffcnrvqtravp	360
QY	361	ffestdiefvlggtamgsfshcyilsalamgypnvpvpslysrfaaqslsvctlmvglltggw	420
Db	361	ffestdiefvlggtamgsfshcyilsalamgypnvpvpslysrfaaqslsvctlmvglltggw	420
QY	421	pvviehfvdkpsil	434
Db	421	pvviehfvdkpsil	434

RESULT 3  
ABB64784  
ID ABB64784 standard; Protein; 458 AA.  
XX  
XX ABB64784;  
XX  
XX 26-MAR-2002 (first entry)  
XX  
XX  
XX  
XX Drosophila melanogaster polypeptide SEQ ID NO 21144.  
XX  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical..  
KW

XX	Drosophila melanogaster.
XX	WO200171042-A2.
XX	27-SEP-2001.
XX	23-MAR-2001; 2001WO-US09231.
XX	23-MAR-2000; 2000US-191637P.
PR	11-JUL-2000; 2000US-0614150.
XX	(PEKE ) PE CORP NY.
XX	Venter JC, Adams M, Li PWD, Myers EW;
PI	WPI: 2001-656860/75.
XX	N-PSDB; ABL08887.
DR	
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more
PT	genes from Drosophila and for elucidating cell signalling and cell-cell
PT	interactions -
XX	
PS	Disclosure; SEQ ID NO 21144; 21pp + Sequence Listing; English.
XX	
CC	The invention relates to an isolated nucleic acid detection reagent
CC	capable of detecting 1000 or more genes from Drosophila. The invention is
CC	useful in developmental biology and in elucidating cell signalling and
CC	cell-cell interactions in higher eukaryotes for the development of
CC	insecticides, therapeutics and pharmaceutical drugs. The invention
CC	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC	sequences (ABL01840-ABL16175) and the encoded proteins
CC	(ABBS57737-ABBY2072)

CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published\\_pat\\_sequences](ftp.wipo.int/pub/published_pat_sequences).

RESULT	4	
AA58286		
ID	AA58286 standard; Protein; 475 AA.	
XX		
XX		
XX	AA58286;	
XX		
DT	19-JUN-2000 (first entry)	
XX		
DE	Rat ENT1 receptor SEQ ID NO:5.	
XX		
KW	Rat; ENT1; ENT1; immunosuppressant	
KW	cytostatic; hypotensive; antiinflam	
KW	nucleoside transporter; ischaemia;	
KW	organ transplant rejection; renitis	
KW	analgesic; blood platelet coagulat	
XX		
OS	Rattus sp.	
XX		
PN	W0200012550-A1.	
XX		
PD	09-MAR-2000.	
XX		
PF	26-AUG-1999; 99WO-JP04502.	
XX		
PR	27-AUG-1998; 98JP-0241248.	
XX		
PA	(KYOW ) KYOWA HAKKO KOGYO KK.	
XX		
PI	Miyaji H, Mimura H, Kambe M, Nak	
XX		
DR	WPI; 2000-256588/22.	
DR	N-PSDB; AAA08030.	
XX		





Db 359 rwlpslvlarlvfvpdlillcn--ikpryltwvfhdawffmaafafngylaslcmc 416  
Qy 389 YTPNVVPSHYSRFAAQLSVCTLMWGLLTGGLPVPIEHFV 428  
Db 417 fgpkkvpaetaetagaiaffclglalgalgavfsifraiv 456

RESULT 6  
AAB15520  
ID AAB15520 standard; Protein; 456 AA.  
AC AAB15520;  
DT 14-FEB-2001 (first entry)  
DE Human ENT1 protein.  
KW Antiviral; antitumour; NBMPR-iENTP; nitrobenzylmercaptapurineriboside;  
KW insensitive, equilibrative nucleoside transporter protein; cancer;  
KW plasma membrane; facilitated diffusion; gene therapy; primer; probe.  
OS Homo sapiens.  
XX US6130065-A.  
XX 10-OCT-2000.  
XX 09-APR-1998; 98US-0058389.  
XX 11-APR-1997; 97US-0043659.  
XX (SUUD-) ST JUDE CHILDREN'S RES HOSPITAL.  
XX Crawford CR, Belt JA;  
XX WPI; 2000-637839/61.  
XX New nucleic acid encoding nucleoside transporter, useful for drug  
XX screening to identify antitumour and antiviral agents and for gene  
XX therapy -  
XX Disclosure; Fig 2A-B; 55pp; English.

The invention relates to the isolation of a gene encoding a NBMPR-iENTP  
(nitrobenzylmercaptapurineriboside-insensitive, equilibrative nucleoside  
transporter protein) from HeLa cells. The protein transports nucleosides  
across the plasma membrane by a facilitated diffusion process and is  
involved in the proliferative response. The transporter is useful in  
screening to identify natural nucleoside permeants and/or their  
inhibitors or analogues, potential therapeutic agents, also for studies  
on protein structure and mechanism. Cells that express the iENTP as the  
only transporter protein are used for drug screening (especially to  
identify antitumour and antiviral nucleoside analogues), in chemotherapy  
of cancer and for selective expression of heterologous genes for gene  
therapy. The cDNA is used for recombinant expression of iENTP and as  
a source of oligonucleotides (diagnostic primers and probes, ribozymes  
and antisense sequences). Antibodies raised against iENTP are used for  
detection of the protein by usual immunoassays and as (antagonists of  
iENTP activity. This sequence represents the human ENT1 protein and  
is used for comparison with the human NBMPR-iENTP protein (AAB15517).

XX Sequence 456 AA;  
Query Match 15.08; Score 343.5; DB 21; Length 456;  
Best Local Similarity 24.38; Pred. No. 1.1e-26;  
Matches 112; Conservative 77; Mismatches 212; Indels 59; Gaps 13;

Qy 18 PRDKYNIYWLIVLGVGGLLPWNWFTIAPEYVYVWFKPDGVE---TWYSKEFMGSLT 74  
Db 7 pdkrykavwlffmglgltlpwnffmt-atqyftnrldmsqnsvlvtalskdaqasaa 65

Qy 75 IGSQLP--NASINVENLELIAGPLIYRVFA-----PVCNFIYV-----LTIILI 117  
Db 66 paaplpernsisaifnnvmticamlplllfytlosflhqrpqsvrlgslvaillvfl 125  
Qy 118 LVIVLEPTEDSMWFFWVTLGMATSIINFSNGLYENSVYGVGGDFPHFYIGALLIGNNICG 177  
Db 126 tailvkqldalp-ffvitmikivlinsfgailgqslfglagllpasytapimsqgglag 184  
Qy 178 LLITVVKIGVTVFLNDEPKLVAIVVFGISLVILLVCAIALPFITKQDFVHHQ----- 231  
Db 185 ffasvami-caiasgselsesafyftacaviliitiicyigpriefryyqqkklsepp 243  
Qy 232 -----KGMEIREKAETDRPSPILWTT-----FTNCYGOLENVWFCAV 271  
Db 244 gegetklldiskgeepragkeesgvsvnsqptneshsikaillknslvafsvcfift 303  
Qy 272 LTIFPVMVTVTRGDSGFNLKIMSENDEIYLLTSFIVNLFPAAGISIVASKIHP--TP 329  
Db 304 igmpavtvevkssiag-----stweryfipvscfltnifdwigrsitavfmpgkds 358  
Qy 330 RYLKFAILLRALFIPFFFCNRYVOTRAY-PVFFESTDIFVIGGTAMSFSGHYLSALAMG 388  
Db 359 rwlpslvlarlvfvpdlillcn--ikpryltwvfhdawffmaafafngylaslcmc 416  
Qy 389 YTPNVVPSHYSRFAAQLSVCTLMWGLLTGGLPVPIEHFV 428  
Db 417 fgpkkvpaetaetagaiaffclglalgalgavfsifraiv 456

RESULT 7

AAY82285  
ID AAY82285 standard; Protein; 475 AA.  
XX AC AAY82285;  
XX DT 19-JUN-2000 (first entry)  
XX DE Human ENT1 receptor SEQ ID NO:1.  
XX KW Human; ENT1; ENT1; immunosuppressant; vasotropic; thrombolytic;  
KW cytostatic; hypotensive; antiinflammatory; analgesic; anticoagulant;  
KW nucleoside transporter; ischaemia; cerebral embolism; malignant tumour;  
KW organ transplant rejection; renitis; pancreatitis; hypertension;  
KW analgesic; blood platelet coagulation inhibitor.  
XX OS Homo sapiens.  
XX PN WO200012550-A1.  
XX PD 09-MAR-2000.  
XX PF 26-AUG-1999; 99WO-JP04602.  
XX PR 27-AUG-1998; 98JP-0241248.  
XX PA (KYOW ) KYOWA HAKKO KOGYO KK.  
XX PL Miyaji H, Mimura H, Kambe M, Nakagawa S;  
XX WPI; 2000-256588/22.  
XX DR N-PSDB; AAA08027.  
XX PT Novel nucleoside transporter polypeptide, useful in drug development  
XX for agents in treating e.g. ischemia, cerebral embolism, rejection in  
XX organ transplant, malignant tumors, renitis, pancreatitis and  
XX hypertension -  
XX PS Claim 1; Page 70-74; 94pp; Japanese.  
XX CC The present invention describes human and rat ENT1 receptors (ENTR1),  
XX with some amino-acids deleted, substituted or added, but which retain  
XX nucleoside transport activity, and can be used in drug development for  
XX agents in treating e.g. ischaemia, cerebral embolism, rejection in organ



XX DE Equilibrative nucleoside transport protein (iENTP).

XX KW Equilibrative nucleoside transport protein; iENTP; NBMPR; transport;

XX KW nitrobenzylmercaptapurine riboside; antiviral; antitumour; screening;

XX KW inhibitor; drug; knockout mice; cancer; chemotherapy; gene therapy;

XX KW adenosine deaminase; factor VIII.

XX OS Homo sapiens.

XX PH Location/Qualifiers

FT Domain 13..27

FT /note= "transmembrane domain TM1"

FT Domain 71..93

FT /note= "transmembrane domain TM2"

FT Domain 100..117

FT /note= "transmembrane domain TM3"

FT Domain 125..144

FT /note= "transmembrane domain TM4"

FT Domain 162..185

FT /note= "transmembrane domain TM5"

FT Domain 194..214

FT /note= "transmembrane domain TM6"

FT Domain 293..312

FT /note= "transmembrane domain TM7"

FT Domain 326..341

FT /note= "transmembrane domain TM8"

FT Domain 361..379

FT /note= "transmembrane domain TM9"

FT Domain 393..413

FT /note= "transmembrane domain TM10"

FT Domain 432..452

FT /note= "transmembrane domain TM11"

XX WO9846749-A1.

XX PD 22-OCT-1998.

XX PF 10-APR-1998; 98WO-US07283.

XX PR 09-APR-1998; 98US-0058389.

XX PR 11-APR-1997; 97US-0838845.

XX PA (SJUD-) ST JUDE CHILDREN'S RES HOSPITAL.

XX PI Belt JA, Crawford CR, Patel DH;

XX WPI; 1998-594576/50.

XX N-PSDB; AAV69698.

XX New equilibrative nucleoside transport protein insensitive to

XX nitrobenzylthio-purine riboside - useful for, e.g. identifying

XX specific therapeutic nucleoside analogues and in gene therapy to

XX protect transduced cells against ablative chemotherapy

XX Claim 13; Pages 71-72; 114pp; English.

XX This represents an equilibrative nucleoside transport protein (iENTP),

XX which is insensitive to nitrobenzylmercaptapurine riboside (NBMPR). Cells

XX transformed with a construct containing the iENTP nucleic acid can be

XX used to produce the protein recombinantly. iENTP is used to identify

XX specific ligands (particularly antiviral and antitumour nucleoside

XX analogues that are preferentially transported into cells) and to raise

XX antibodies. Cells in which iENTP provides all available transport

XX activity are used: (a) to identify permeants of iENTP and (b) to screen

XX specific inhibitors of iENTP (potential drugs). Fragments of the iENTP

XX nucleic acid are used, as probes, primers, antisense molecules, and

XX ribozymes for therapy or diagnosis, and knockout mice in which both

XX alleles encoding iENTP contain an inactivating defect are also useful for

XX drug screening. Cells that have been transduced with iENTP nucleic acid

XX ex vivo are used particularly, for cancer chemotherapy. Vectors in which

XX the iENTP gene is linked to a heterologous gene (e.g. encoding adenosine

XX deaminase and factor VIII) can be used for gene therapy.

XX SQ Sequence 456 AA;

Query Match 14.2%; Score 326.5; DB 19; Length 456;

Best Local Similarity 22.9%; Pred. No. 6.3e-25;

Matches 106; Conservative 71; Mismatches 200; Indels 85; Gaps 11;

QY 18 PRDKYNIIVWLIVLGVGLLPNMFIITIAPEY-----VNYWFKPDGYE 62

DB 7 prdsynlvgisffliglgtllpwnffitaipyfqarlgaganstarilstnhtgpedaf- 65

QY 63 TWYSKEFMGSLTIGSQLPNASINVFNLFLIIAGPLIYRVFAPVCFNIVNLTIILIVYL 122

DB 66 -----nfnwnvtllsqpllllftllnsflyqcvpetvrilg----sllaillfaltaal 116

QY 123 EPTEDSMSEFFFWTLCMATSFNSNGLYENSVYGVGGDPHPTHYIGALLIGNICGLLITV 182

DB 117 vkvdmspgpffsittmasvcfinsfsaviqgsifglgtmpstystlffsggqlagifaa 176

QY 183 VKI-----GVTYFLNDEPKLVAIVF---GISLVILLVCAIALFFITKQDFY---HYHHQ 231

DB 177 amlslmasgv-----daetsalgylfipcvgilmsivcylslphlkkfaryylankssa 230

QY 232 KGMEIREKA-----ETDRPSPSILMTTFTNC 257

DB 231 qaqlletkaellqsdengipsspqkvallldldlekepesepepqpqkpsvftvfqki 290

QY 258 YQOLFNVWFCEFAVTLTIPTVMVTTRGDSGFLNKIMSENDEIYTLTSLFVFNLFAAIG 317

DB 291 witaiclvlvftvtlsvipaitamvtsts-----pgkwsqffpiccflflnmwdwig 344

QY 318 SIVASKIHWP--TPRYLKFALILRALFIPFFFCNVRVOTRAYPVFFESTDIFVIGGIAM 375

DB 345 rsitsyflwpdedsrllpllvclrfvlvplfmlchvpqrsr-lpilfpqdayfitmlif 403

QY 376 SFSHGYSALAMGYTPNVVPSHYSRFAAQLSVCTLMVGLLTG 417

DB 404 avsnngylvsiltnciaprqvlppherevagalmftflalglscg 445

RESULT 10

AAW69558

ID AAW69558 standard; Protein; 456 AA.

XX AC AAW69558;

XX DT 13-OCT-1998 (first entry)

XX DE Human equilibrative nucleoside transporter 2.

XX KW Human; equilibrative nucleoside transporter; hENT1; hENT2; rENT1;

XX KW rENT2; coronary; cerebrovascular anoxia; viral infection; cancer.

XX OS Homo sapiens.

XX PN WO9829437-A2.

XX PD 09-JUL-1998.

XX PF 30-DEC-1997; 97WO-IB01657.

XX PR 03-NOV-1997; 97US-0064004.

XX PR 30-DEC-1996; 96US-0034083.

XX PA (UVAL-) UNIV ALBERTA.

XX PA (UYLE-) UNIV LEEDS.

XX PI Baldwin SA, Cass CE, Young JD;

XX WPI; 1998-388035/33.

XX N-PSDB; AAV40279.

PT Newly isolated equilibrative nucleoside transporter protein(s) and  
PT gene(s) - used to develop products for treating disorder(s)  
PT associated with the transporter(s) and for use with nucleoside  
PT drug(s)  
XX  
XX Claim 5; Fig 20; 97pp; English.  
XX  
XX The present sequence represents a substantially purified equilibrative  
XX nucleoside transporter (ENT), human ENT2 (iENT2). ENTs can transport a  
XX variety of purines and pyrimidines, including adenosine, uridine,  
XX guanosine, inosine, formycin B, tubercidin, and thymidine. ENTs are  
XX bidirectional, they transport a suitable permeant both into and out of  
XX cells. ENTs can be used as a tool for the development of new nucleoside  
XX drugs. Products from the present invention can be used for treating a  
XX subject having a disorder associated with an ENT. They can also be used  
XX with nucleoside drugs in the treatment of e.g. coronary or  
XX cerebrovascular anoxia, viral infection or cancer. The products (e.g.  
XX antibodies and oligonucleotides hybridising to nucleic acid sequences  
XX encoding ENTs) can also be used for detection and drug screening.  
XX  
XX Sequence 456 AA;  
XX  
XX  
XX Query Match 14.2%; Score 326.5; DB 19; Length 456;  
XX Best Local Similarity 22.9%; Pred. No. 6.3e-25;  
XX Matches 106; Conservative 71; Mismatches 200; Indels 85; Gaps 11;  
XX  
XX QY 18 PRDKYNIYVWLILVFGVLLPWNFFIIIAPEYY-----VNWFKPDGVE 62  
XX Db 7 prdsyhlvgisffilglgtllpwnffitaipfqrarlagagnstarilstnhtgpedaf- 65  
XX  
XX QY 63 TWYSKEFGSUTIGSOLPNASINVENFLIITAGPLIYVFAPVCNFINLTIILIVL 122  
XX Db 66 -----nfnnwvllsqplllftlnslfycvpetvrlg----sllaillfaltaal 116  
XX  
XX QY 123 EPTDSMSWFFWVLTGMATSNFNSGLYENSIVYGGDFPHYIIGALLIGNNICLLITV 182  
XX Db 117 vkvdmspgpffsitmasvcfinsfsavlgqslfglgtmpstystflfsggqlagifaa 176  
XX  
XX QY 183 VKI-----GVTYFLNDEPKLVAIVF---GISLVLLVCAITAFITTKQDFY---HYHHQ 231  
XX Db 177 amllsmasgv-----daetsalgyfltpcvgilmsivcylslphikfaryylankssqa 230  
XX  
XX QY 232 KGMETREKA-----ETDRSPSILWTTFTNC 257  
XX Db 231 qaqeletkaellqsdengipspqkvaltdldlekepesepepkpgkpsvftvfaki 290  
XX  
XX QY 258 YGQLENWFCFAVLTITPPVMVMTVTRGDSGFLNKIMSENDEIYTLTSLFVNLFAAIG 317  
XX Db 291 witalclvlvftvtlsvfpaitamvtssts-----pgkwsqffnppiccfllfnmdwlg 344  
XX  
XX QY 318 SIVASKIHWP--TPRYLKFAILLRALFTPEFFFCNRYQTRAYPVFFESTDIEVIGGIA 375  
XX Db 345 rsitsyflwpedrsrlplllclrfifvplfmlchvqrsr-lpllfpqdayfifcmlf 403  
XX  
XX QY 376 SFSHGYSALAMGYTPNVPSHYSRFAQLSVCTLMVGLLTG 417  
XX Db 404 avsnqylvslmtclaprqlpherevagalmftflalqlscg 445  
XX  
XX RESULT 11  
XX ID AAB15517  
XX AC AAB15517;  
XX DT 14-FEB-2001 (first entry)  
XX DE Human NBMPR-iENTP protein.  
XX KW Antiviral; anticancer; NBMPR-iENTP; nitrobenzylmercaptopurineriboside;  
XX insensitve, equilibrative nucleoside transporter protein; cancer;  
XX plasma membrane; facilitated diffusion; gene therapy; primer; probe.

XX Homo sapiens.  
XX OS US6130065-A.  
XX PN 10-OCT-2000.  
XX PD 09-APR-1998; 98US-0058389.  
XX PF 11-APR-1997; 97US-0043659.  
XX PR (STUD-) ST JUDE CHILDREN'S RES HOSPITAL.  
XX PA Crawford CR, Belt JA;  
XX PI WPI; 2000-637839/61.  
XX DR N-PSDB; AAA95722.  
XX DR New nucleic acid encoding nucleoside transporter, useful for drug  
XX screening to identify antitumour and antiviral agents and for gene  
XX therapy -  
XX  
XX Claim 1; Fig 2A-B; 55pp; English.  
XX  
XX This sequence represents nitrobenzylmercaptopurineriboside-insensitive,  
XX equilibrative nucleoside transporter protein (NBMPR-iENTP) isolated from  
XX HeLa cells. The protein transports nucleosides across the plasma membrane  
XX by a facilitated diffusion process and is involved in the proliferative  
XX response. The transporter is useful in screening to identify natural  
XX nucleoside permeants and/or their inhibitors or analogues, potential  
XX therapeutic agents, also for studies on protein structure and mechanism.  
XX Cells that express the iENTP as the only transporter protein are used  
XX for drug screening (especially to identify antitumour and antiviral  
XX nucleoside analogues), in chemotherapy of cancer and for selective  
XX expression of heterologous genes for gene therapy. The cDNA is used  
XX for recombinant expression of iENTP and as a source of oligonucleotides  
XX (diagnostic primers and probes, ribozymes and antisense sequences).  
XX Antibodies raised against iENTP are used for detection of the protein  
XX by usual immunoassays and as (ant)agonists of iENTP activity.  
XX  
XX Sequence 456 AA;  
XX  
XX  
XX Query Match 14.2%; Score 326.5; DB 21; Length 456;  
XX Best Local Similarity 22.9%; Pred. No. 6.3e-25;  
XX Matches 106; Conservative 71; Mismatches 200; Indels 85; Gaps 11;  
XX  
XX QY 18 PRDKYNIYVWLILVFGVLLPWNFFIIIAPEYY-----VNWFKPDGVE 62  
XX Db 7 prdsyhlvgisffilglgtllpwnffitaipfqrarlagagnstarilstnhtgpedaf- 65  
XX  
XX QY 63 TWYSKEFGSUTIGSOLPNASINVENFLIITAGPLIYVFAPVCNFINLTIILIVL 122  
XX Db 66 -----nfnnwvllsqplllftlnslfycvpetvrlg----sllaillfaltaal 116  
XX  
XX QY 123 EPTDSMSWFFWVLTGMATSNFNSGLYENSIVYGGDFPHYIIGALLIGNNICLLITV 182  
XX Db 117 vkvdmspgpffsitmasvcfinsfsavlgqslfglgtmpstystflfsggqlagifaa 176  
XX  
XX QY 183 VKI-----GVTYFLNDEPKLVAIVF---GISLVLLVCAITAFITTKQDFY---HYHHQ 231  
XX Db 177 amllsmasgv-----daetsalgyfltpcvgilmsivcylslphikfaryylankssqa 230  
XX  
XX QY 232 KGMETREKA-----ETDRSPSILWTTFTNC 257  
XX Db 231 qaqeletkaellqsdengipspqkvaltdldlekepesepepkpgkpsvftvfaki 290  
XX  
XX QY 258 YGQLENWFCFAVLTITPPVMVMTVTRGDSGFLNKIMSENDEIYTLTSLFVNLFAAIG 317  
XX Db 291 witalclvlvftvtlsvfpaitamvtssts-----pgkwsqffnppiccfllfnmdwlg 344  
XX  
XX QY 318 SIVASKIHWP--TPRYLKFAILLRALFTPEFFFCNRYQTRAYPVFFESTDIEVIGGIA 375  
XX Db 345 rsitsyflwpedrsrlplllclrfifvplfmlchvqrsr-lpllfpqdayfifcmlf 403  
XX  
XX QY 376 SFSHGYSALAMGYTPNVPSHYSRFAQLSVCTLMVGLLTG 417  
XX Db 404 avsnqylvslmtclaprqlpherevagalmftflalqlscg 445  
XX  
XX RESULT 11  
XX ID AAB15517  
XX AC AAB15517;  
XX DT 14-FEB-2001 (first entry)  
XX DE Human NBMPR-iENTP protein.  
XX KW Antiviral; anticancer; NBMPR-iENTP; nitrobenzylmercaptopurineriboside;  
XX insensitve, equilibrative nucleoside transporter protein; cancer;  
XX plasma membrane; facilitated diffusion; gene therapy; primer; probe.













**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 23, 2002, 17:35:27 ; Search time 3087.27 Seconds  
(without alignments)  
6068.071 Million cell updates/sec

Title: US-08-816-011f-36  
Perfect score: 1388  
Sequence: 1 atggttaataatcagcagatc.....ttattaaaaaaaaaaaaa 1388

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues  
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estcov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_estl:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_gss:\*  
13: em\_gss\_hum:\*  
14: em\_gss\_inv:\*  
15: em\_gss\_pln:\*  
16: em\_gss\_vrt:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	96.2	6.9	551	10	BJ119282	BJ119282 BJ119282
C 2	96.2	6.9	669	10	BJ147020	BJ147020 BJ147020
C 3	86.6	6.2	533	9	AU222343	AU222343 AU222343
4	77.6	5.6	332	9	AU201199	AU201199 AU201199
C 5	75.6	5.4	710	9	AU216992	AU216992 AU216992
6	75.2	5.4	648	9	AW453425	AW453425 SWOV3MCA
7	73.8	5.3	327	9	AU109268	AU109268 AU109268
8	73.8	5.3	642	9	AU1082929	AU1082929 SWAMCAG12
9	70.8	5.1	495	9	AU209436	AU209436 AU209436
10	70.2	5.1	360	10	CI3829	CI3829 CI3829 Yuj1
C 11	67.8	4.9	300	9	AU113471	AU113471 AU113471
C 12	57.6	4.1	300	9	AU115402	AU115402 AU115402
C 13	57.4	4.1	300	9	AU114995	AU114995 AU114995
14	56.2	4.0	997	12	CNS005TE	AU060767 Drosophila
15	54.4	3.9	258	9	AA406898	AA406898 MBACF27F0
C 16	54.4	3.9	543	10	BI501984	BI501984 rm07f05.Y
17	53	3.8	656	9	AW409482	AW409482 SWOV3MCA

18	50.4	3.6	345	9	AU109592	AU109592 AU109592
19	49	3.5	606	9	AA948920	AA948920 LD27618.5
C 20	48.8	3.5	901	12	CNS0760F	AL430789 clone XBA
21	48.2	3.5	1201	12	CNS0010J	AL054622 Drosophila
22	47.6	3.4	274	9	AA842090	AA842090 MBACF380
23	47	3.4	228	9	AI784877	AI784877 SWAMCAG40
24	46.8	3.4	729	9	AI239052	AI239052 GH15222.5
25	46.2	3.3	360	10	R03450	R03450 pk06g03.r1
26	46.2	3.3	373	10	C43216	C43216 C43216 Yuj1
27	45.8	3.3	374	9	BB181657	BB181657 BB181657
28	45.6	3.3	350	9	AU111089	AU111089 AU111089
29	45.6	3.3	351	9	AU111161	AU111161 AU111161
30	45.6	3.3	738	12	AQ842781	AQ842781 CpG1299A
C 31	45.6	3.3	759	12	AQ254298	AQ254298 CpG0751B
32	45	3.2	375	10	C44886	C44886 C44886 Yuj1
33	44.6	3.2	340	9	AU111560	AU111560 AU111560
34	44.6	3.2	340	9	AU111734	AU111734 AU111734
C 35	44.6	3.2	510	12	AQ254261	AQ254261 CpG0732A
C 36	44.4	3.2	429	9	AW829621	AW829621 ra42d09.y
C 37	44.4	3.2	448	9	AW829541	AW829541 ra41d09.y
C 38	44.2	3.2	333	9	AI431074	AI431074 mj58g09.x
C 39	44.2	3.2	441	10	BF459052	BF459052 UI-M-B21-
40	44.2	3.2	442	9	BB750597	BB750597 BB750597
41	44.2	3.2	459	9	AI595433	AI595433 mj58g09.y
42	44.2	3.2	475	9	AA058117	AA058117 mj58g09.r
C 43	44.2	3.2	566	9	AW333515	AW333515 S22G3 AGS
C 44	44.2	3.2	626	9	AW048063	AW048063 UI-M-BH1-
45	44.2	3.2	649	9	BB169820	BB169820 BB169820

ALIGNMENTS

RESULT 1  
BJ119282  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
BASE COUNT  
ORIGIN  
Query Match

551 bp mRNA linear EST 23-JAN-2002  
unpublished oligo-capped cDNA library, C. elegans L1 stage  
Caenorhabditis elegans cDNA clone yk1241g07 5', mRNA sequence.  
BJ119282.1 GI:18279408  
Caenorhabditis elegans.  
Caenorhabditis elegans  
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea  
; Rhabditidae; Peloderinae; Caenorhabditis.  
1 (bases 1 to 551)  
Kohara, F., Shin-I, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.  
and Sugano, S.  
A complementary view of the C.elegans genome  
Unpublished (2002)  
Contact: Tadasu Shin-i  
Center For Genetic Resource Information  
National Institute of Genetics  
1111 Yata, Mishima, Shizuoka 411-8540, Japan  
Tel: 81-559-81-6856  
Fax: 81-559-81-6855  
Email: tshini@genes.nig.ac.jp.  
Location/Qualifiers  
I. .551  
/organism="Caenorhabditis elegans"  
/strain="N2"  
/db\_xref="taxon:6239"  
/clone="yk1241g07"  
/clone\_lib="unpublished oligo-capped cDNA library, C.  
elegans L1 stage"  
/sex="hermaphrodite"  
/tissue\_type="whole animal"  
/dev\_stage="L1"

117 a 152 c 94 g 187 t 1 others  
6.9%; Score 96.2; DB 10; Length 551;

Best Local Similarity 53.6%; Pred. No. 1.8e-09;

Matches 200; Conservative 0; Mismatches 173; Indels 0; Gaps 0;

```
Qy 911 tcacaagtttcctgcgttcacattgttcgctgcgagtgatccatagttgcttccaaga 970
Db 27 TGACCACATTCCTTCAATTCAACGTTGTCGTTTATTGGATCAATTTGCGCGGGAAGAA 86
Qy 971 ttcactggcgacacaccccgcttaccataaatttgccataaattcgtgctcttttcattc 1030
Db 87 AGCAATGGCTCGCGGACAAAGCTCTGGATTCCAGTCTACCTCCGTCCTCTATATTC 146
Qy 1031 catttcttcttcgcaactatcgtgtccagacggtgcttattcctgttttttttgagt 1090
Db 147 CATTTCTTCATCTCTCTGCAACTATCTCCCTGAGACCGGTTCACTTCCCGTCTTTTTCGAAT 206
Qy 1091 ctactgacatttttgattggtggaattggcattgctctttttcacatgatacctcagcg 1150
Db 207 CCACCTGGCTTTTCGTCATTATCGCGGCTCGATGAGCTTTGGAGTGGATATTCCTCGG 266
Qy 1151 ctctggcaatgggatacacctccaaacgtgctgccattctcactactcaagatttgcgctc 1210
Db 267 GACTCGCATGATGTACACCTCGAAGACTGTTGATCCATCGAAGGCTCAGGTCGCTGGAA 326
Qy 1211 agctttccggttgcaactcttattggtgaccttcacacggtgacctggtgcccgtttgta 1270
Db 327 TGATGGCGGATTTCTTCCTCATCTCTGGAATTGCTCTGCTGTGATCTTCACAATGGTCA 386
Qy 1271 ttgagcaacttcgt 1283
Db 387 TCAAGATGTCGT 399
```

RESULT 2

BJ147020/c

LOCUS

DEFINITION BJ147020 unpublished oligo-capped cDNA library, C. elegans L1 stage

Caenorhabditis elegans cDNA clone yk1241g07 3', mRNA sequence.

ACCESSION BJ147020

VERSION

KEYWORDS

SOURCE

ORGANISM

Caenorhabditis elegans.

Caenorhabditis elegans

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea

; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE 1 (bases 1 to 669)

AUTHORS Kohara,Y., Shin-I,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.

and Sugano,S.

TITLE A complementary view of the C.elegans genome

JOURNAL Unpublished (2002)

COMMENT Contact: Tadasu Shin-i

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan

Tel: 81-559-81-6856

Fax: 81-559-81-6855

Email: tshini@genes.nig.ac.jp.

FEATURES

Location/Qualifiers

1..669

/organism="Caenorhabditis elegans"

/strain="N2"

/db\_xref="taxon:6239"

/clone="yk1241g07"

/clone\_lib="unpublished oligo-capped cDNA library, C.

elegans L1 stage"

/sex="hermaphrodite"

/tissue\_type="whole animal"

/dev\_stage="L1"

BASE COUNT 240 a 106 c 170 g 151 t 2 others

ORIGIN

Query Match

Best Local Similarity 53.6%; Pred. No. 1.7e-09;

Matches 173; Indels 0; Gaps 0;

Matches 200; Conservative 0; Mismatches 173; Indels 0; Gaps 0;

```
Qy 911 tcacaagtttcctgcgttcacattgttcgctgcgagtgatccatagttgcttccaaga 970
Db 659 TGACCACATTCCTTCAATTCAACGTTGTCGTTTATTGGATCAATTTGCGCGGGAAGAA 600
Qy 971 ttcactggcgacacaccccgcttaccataaatttgccataaattcgtgctcttttcattc 1030
Db 599 AGCAATGGCTCGCGGACAAAGCTCTGGATTCCAGTCTACCTCCGTCCTCTATATTC 540
Qy 1031 catttcttcttcgcaactatcgtgtccagacggtgcttattcctgttttttttgagt 1090
Db 539 CATTTCTTCATCTCTCTGCAACTATCTCCCTGAGACCGGTTCACTTCCCGTCTTTTTCGAAT 480
Qy 1091 ctactgacatttttgattggtggaattggcattgctctttttcacatgatacctcagcg 1150
Db 479 CCACCTGGCTTTTCGTCATTATCGCGGCTCGATGAGCTTTGGAGTGGATATTCCTCGG 420
Qy 1151 ctctggcaatgggatacacctccaaacgtgctgccattctcactactcaagatttgcgctc 1210
Db 419 GACTCGCATGATGTACACCTCGAAGACTGTTGATCCATCGAAGGCTCAGGTCGCTGGAA 360
Qy 1211 agctttccggttgcaactcttattggtgaccttcacacggtgacctggtgcccgtttgta 1270
Db 359 TGATGGCGGATTTCTTCCTCATCTCTGGAATTGCTCTGCTGTGATCTTCACAATGGTCA 300
Qy 1271 ttgagcaacttcgt 1283
Db 299 TCAAGATGTCGT 287
```

RESULT 3

AU222343/c

LOCUS

DEFINITION AU222343 unpublished oligo-capped cDNA library, stage L1

Caenorhabditis elegans cDNA clone yk1015b04 3', mRNA sequence.

ACCESSION AU222343

VERSION

KEYWORDS

SOURCE

ORGANISM

Caenorhabditis elegans.

Caenorhabditis elegans

Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea

; Rhabditidae; Peloderinae; Caenorhabditis.

REFERENCE 1 (bases 1 to 533)

AUTHORS Kohara,Y., Shin-I,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.

and Sugano,S.

TITLE A complementary view of the C.elegans genome

JOURNAL Unpublished (2001)

COMMENT Contact: Yuji Kohara

Genome Biology Lab.

National Institute of Genetics

Yata 1111, Mishima, Shizuoka 411, Japan

Tel: 81-559-81-6854

Fax: 81-559-81-6855

Email: ykohara@lab.nig.ac.jp.

FEATURES

Location/Qualifiers

1..533

/organism="Caenorhabditis elegans"

/strain="N2"

/db\_xref="taxon:6239"

/clone="yk1015b04"

/clone\_lib="unpublished oligo-capped cDNA library, stage

L1"

/sex="Hermaphrodite"

/tissue\_type="whole animal"

/dev\_stage="L1"

BASE COUNT 191 a 101 c 100 g 141 t

ORIGIN

Query Match

Best Local Similarity 49.3%; Pred. No. 1.5e-07;

Matches 255; Conservative 0; Mismatches 259; Indels 3; Gaps 1;

```
Qy 808 gttactctcaaatcttccctgttatgatgaccgtttacacactcgtggagattcccgcttc 867
Db 514 GTTTCCTCTCTCATCTTTCCCTGCAATGACTGATAGTGTGACTCTGATCCTACAAAT 455
Qy 868 ctaaacaaaaatagtctgaaacagatgaaatctacacatttgctcacaagtttctctcgc 927
Db 454 GGAATAAATCTGTTGTTTGGAGATAGTTATTTCTCCG---GATAAATACACTTCTCTCAAT 398
Qy 928 ttcaattgttcgcgtgagatgacccatagttgcttccaaagattcactggccgacaccc 987
Db 397 TTCAATTTATTGATGATGGATGGATCATCTCTTGCCAAATATGTTCAATCCCATCAGAG 338
Qy 988 cgttaccctcaaatggcataatcttgctgctctcttccattccattctctctctcgc 1047
Db 337 AAATATTATTGATGGAGTGGCTCTTCGAACCGTATTCATTCATTTATCTTTCTGCT 278
Qy 1048 aactatcgtccagacgcgtgcttatcctcgtttctcttgtagtctactgacattttgtg 1107
Db 277 AATTATCTCCAAATPACTAGAAGATGGCGCTGTTGGTTTAAAGAATGAATGGTGGTCACT 218
Qy 1108 attggtgaattgcattctcttttccatcgtatcaccctcagcgtctcgtgcaatggatac 1167
Db 217 ATTGATGATACAAATGGCAATTTACTTGTGTTATATGAGCAGTATTAGCGCTGATTTAT 158
Qy 1168 actcaaacgctgctgccatctcactcactcaagattgcccgtcagcttccgtttgcact 1227
Db 157 ACACCAAGCAAGTCCAGCTAGATATCAGAAGCTAAGTGGAAATGCTTGTCATCAATTTTC 98
Qy 1228 cttatggtgctcctctcaccggtggtcgtggtgcccgttgattgagcactctcgtgac 1287
Db 97 CTAATGCTCGGAATTTCTATCGGAGTCCAGTACACCAATTCGTGCATGGCGGTGGAC 38
Qy 1288 aagcaagtatctataaatattatatacattagatt 1324
Db 37 TCGATAGGAAGTAGAAGACATGACTGTATAAATTT 1

RESULT 4
LOCUS AU201199 332 bp mRNA linear EST 17-JUL-2001
DEFINITION AU201199 unpublished oligo-capped cDNA library, stage L4
Caenorhabditis elegans cDNA clone yk789e10 5', mRNA sequence.
ACCESSION AU201199
VERSION AU201199.1 GI:14829871
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 332)
AUTHORS Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
and Sugano,S.
TITLE A complementary view of the C.elegans genome
JOURNAL Unpublished (2001)
COMMENT Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers
1. 332
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk789e10"
/clone_lib="unpublished oligo-capped cDNA library, stage
L4"
/sex="Hermaphrodite"
/tissue_type="whole animal"
/dev_stage="L4"

FEATURES
source
1. 332
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk789e10"
/clone_lib="unpublished oligo-capped cDNA library, stage
L4"
/sex="Hermaphrodite"
/tissue_type="whole animal"
/dev_stage="L4"

BASE COUNT 251 a 103 c 181 g 172 t 3 others
ORIGIN

Query Match 5.4%; Score 75.6; DB 9; Length 710;
Best Local Similarity 51.6%; Pred. No. 2.4e-05;
Matches 199; Conservative 0; Mismatches 184; Indels 3; Gaps 1;

Qy 949 ggatccatagttgcttccaaagattcactggccgacacccggttaacctaaatttgcata 1008
Db 709 GGATCAATCGTCGCGGGAGAAAGCAATGGCCGCCGCCGCAACAAGCTCTGGATCCAGTC 650
Qy 1009 actctgctgctcttcttccattccattctctctctctcgaactatcgttccagacggt 1068
Db 649 TACCTCCGTCCTCTACATCCCATCTTCATCTTCTGCAACTCCCTCCCTGAGACCCGC 590
Qy 1069 gcttatcctgtttcttcttgagctactgacatttttgtagtggggaattgccaatgctct 1128
```

```
Qy 808 gttactctcaaatcttccctgttatgatgaccgtttacacactcgtggagattcccgcttc 867
Db 514 GTTTCCTCTCTCATCTTTCCCTGCAATGACTGATAGTGTGACTCTGATCCTACAAAT 455
Qy 868 ctaaacaaaaatagtctgaaacagatgaaatctacacatttgctcacaagtttctctcgc 927
Db 454 GGAATAAATCTGTTGTTTGGAGATAGTTATTTCTCCG---GATAAATACACTTCTCTCAAT 398
Qy 928 ttcaattgttcgcgtgagatgacccatagttgcttccaaagattcactggccgacaccc 987
Db 397 TTCAATTTATTGATGATGGATGGATCATCTCTTGCCAAATATGTTCAATCCCATCAGAG 338
Qy 988 cgttaccctcaaatggcataatcttgctgctctcttccattccattctctctcgc 1047
Db 337 AAATATTATTGATGGAGTGGCTCTTCGAACCGTATTCATTCATTTATCTTTCTGCT 278
Qy 1048 aactatcgtccagacgcgtgcttatcctcgtttctcttgtagtctactgacattttgtg 1107
Db 277 AATTATCTCCAAATPACTAGAAGATGGCGCTGTTGGTTTAAAGAATGAATGGTGGTCACT 218
Qy 1108 attggtgaattgcattctcttttccatcgtatcaccctcagcgtctcgtgcaatggatac 1167
Db 217 ATTGATGATACAAATGGCAATTTACTTGTGTTATATGAGCAGTATTAGCGCTGATTTAT 158
Qy 1168 actcaaacgctgctgccatctcactcactcaagattgcccgtcagcttccgtttgcact 1227
Db 157 ACACCAAGCAAGTCCAGCTAGATATCAGAAGCTAAGTGGAAATGCTTGTCATCAATTTTC 98
Qy 1228 cttatggtgctcctctcaccggtggtcgtggtgcccgttgattgagcactctcgtgac 1287
Db 97 CTAATGCTCGGAATTTCTATCGGAGTCCAGTACACCAATTCGTGCATGGCGGTGGAC 38
Qy 1288 aagcaagtatctataaatattatatacattagatt 1324
Db 37 TCGATAGGAAGTAGAAGACATGACTGTATAAATTT 1

RESULT 4
LOCUS AU201199 332 bp mRNA linear EST 17-JUL-2001
DEFINITION AU201199 unpublished oligo-capped cDNA library, stage L4
Caenorhabditis elegans cDNA clone yk789e10 5', mRNA sequence.
ACCESSION AU201199
VERSION AU201199
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 332)
AUTHORS Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
and Sugano,S.
TITLE A complementary view of the C.elegans genome
JOURNAL Unpublished (2001)
COMMENT Contact: Yuji Kohara
Genome Biology Lab.
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers
1. 332
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk789e10"
/clone_lib="unpublished oligo-capped cDNA library, stage
L4"
/sex="Hermaphrodite"
/tissue_type="whole animal"
/dev_stage="L4"

FEATURES
source
1. 332
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/clone="yk789e10"
/clone_lib="unpublished oligo-capped cDNA library, stage
L4"
/sex="Hermaphrodite"
/tissue_type="whole animal"
/dev_stage="L4"
```



using reverse transcriptase and oligo(dT) followed by

Yata 1111, Mishima, Shizuoka 411, Japan  
Tel: 81-559-81-6854  
Fax: 81-559-81-6855





```
BASE COUNT      114 a      50 c      59 g      77 t
ORIGIN

Query Match      4.9%; Score 67.8; DB 9; Length 300;
Best Local Similarity 55.8%; Pred. No. 0.001;
Matches 129; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

Qy 1024 ttcattccattcttcttctgcaactatcggtccagacgcgtgcttatccctgttttc 1083
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 299 TTCATTCATCTTCATGTTTCAGCAACTATTACCAACATTCCTGCTACTATGGGAGTATTA 240

Qy 1084 tttagtctactgacattttgtgattggtggaattgccatgtctttttcacatggatcac 1143
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 239 TTCACCAACGAATGGATCTCTCTTTTGGTAACACATATTATGGCATTACACAAGTGGATAT 180

Qy 1144 ctcaagcctctggcaatgggaatacacactccaaacgtcgtgccatctcaactcaagattt 1203
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 179 TCTCATCTTTGGGAATGATGATATACACCAACGCTGCTGCCACCGGAATACTCAAAATTA 120

Qy 1204 gcccgtcagctttccgtttgcaactcttatggttgcccttctcaccggtggc 1254
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 119 GCTGCTCAAGTCTGCATGCTCTCTGCTTCTGCTGTTCTGCTATCTACTGCTGCTGTC 69

RESULT 12
AUI15402/c
LOCUS      AUI15402      300 bp      mRNA      linear      EST 19-OCT-2000
DEFINITION      AUI15402 unpublished oligo-capped cDNA library Caenorhabditis
                  elegans cDNA clone yK732c11 3', mRNA sequence.
ACCESSION      AUI15402
VERSION      AUI15402.1 GI:10928969
KEYWORDS
SOURCE
ORGANISM      Caenorhabditis elegans.
                  Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
REFERENCE      1 (bases 1 to 300)
AUTHORS      Kohara,Y., Shin-i.T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
and Sugano,S.
TITLE      A complementary view of the C. elegans genome
JOURNAL      Unpublished (2000)
COMMENT      Contact: Yuji Kohara
                  Genome Biology Lab.
                  National Institute of Genetics
                  Yata 1111, Mishima, Shizuoka 411, Japan
                  Tel: 81-559-81-6854
                  Fax: 81-559-81-6855
                  Email: ykohara@lab.nig.ac.jp.
FEATURES
source
    1..300
        /organism="Caenorhabditis elegans"
        /strain="N2"
        /db_xref="taxon:6239"
        /clone="yK732c11"
        /clone_lib="unpublished oligo-capped cDNA library"
        /sex="Hermaphrodite"
        /tissue_type="whole animal"
        /dev_stage="varied"

BASE COUNT      93 a      68 c      56 g      83 t
ORIGIN

Query Match      4.1%; Score 57.6; DB 9; Length 300;
Best Local Similarity 52.5%; Pred. No. 0.12;
Matches 126; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

Qy 1015 cgtgctcttttcattccattcttcttctgcaactatcggtccagacgcgtgcttat 1074
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 299 CGAACCGTATTCATTCATCTTTATCTTTTCTGTAATATCGTCCAGATACATAGAGATGG 240

Qy 1075 cctgtttcttctgagtcactgacatttttgtgattggtggaattgcatgtcttttca 1134
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db 239 CCTGTTTGGTTTAAGAATGAATGCTGTTCTACTATTGGATGTACCATAATGGCATTTACT 180
Qy 1135 catgataacctcagcgcctctggcaatggagatacaactccaaacgtcgctgccaactcac 1194
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 179 TGTGTTTATATGACAGTATTAGCGCTGATTTATATACCAACGAAGTCCAGCTAGATAT 120
Qy 1195 tcaagatttcgcgtcagcttccgttttgcactcttattggttgcccttctcaccggtggc 1254
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 119 CAGAAGCTAAGTGAATGCTTGTCATCAATTTTCTAATGCTCGGAATTTCTCATCGGAGTC 60

RESULT 13
AUI14995/c
LOCUS      AUI14995      300 bp      mRNA      linear      EST 19-OCT-2000
DEFINITION      AUI14995 unpublished oligo-capped cDNA library Caenorhabditis
                  elegans cDNA clone yK724f10 3', mRNA sequence.
ACCESSION      AUI14995
VERSION      AUI14995.1 GI:10928562
KEYWORDS
SOURCE
ORGANISM      Caenorhabditis elegans.
                  Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
REFERENCE      1 (bases 1 to 300)
AUTHORS      Kohara,Y., Shin-i.T., Thierry-Mieg,J., Thierry-Mieg,D., Suzuki,Y.
and Sugano,S.
TITLE      A complementary view of the C. elegans genome
JOURNAL      Unpublished (2000)
COMMENT      Contact: Yuji Kohara
                  Genome Biology Lab.
                  National Institute of Genetics
                  Yata 1111, Mishima, Shizuoka 411, Japan
                  Tel: 81-559-81-6854
                  Fax: 81-559-81-6855
                  Email: ykohara@lab.nig.ac.jp.
FEATURES
source
    1..300
        /organism="Caenorhabditis elegans"
        /strain="N2"
        /db_xref="taxon:6239"
        /clone="yK724f10"
        /clone_lib="unpublished oligo-capped cDNA library"
        /sex="Hermaphrodite"
        /tissue_type="whole animal"
        /dev_stage="varied"

BASE COUNT      96 a      67 c      53 g      84 t
ORIGIN

Query Match      4.1%; Score 57.4; DB 9; Length 300;
Best Local Similarity 49.5%; Pred. No. 0.13;
Matches 148; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

Qy 1022 ttttcattccattcttcttcttctgcaactatcggtccagacgcgtgcttatccctgttt 1081
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 300 TATTCATTCCTCAATTTATCTTTTCTGTAATATCGTCCAGATACATAGAGATGGCTGTTT 241

Qy 1082 tctttgagtcactgacatttttgtgattggtggaattgccaatgtctttttcaatgagat 1141
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 240 GGTTTAAGAATGAATGGTGGTTCCTACTATTGGATGTACCATTAATGGCATTTTCTTGTGGTT 181

Qy 1142 acctcagcgcctctggcaatgggatacacactccaaacgtcgtgccatctcaactacaagat 1201
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 180 ATATGACAGCTTTAGCGCTGATTTATACACCAACGAAGTCCAGCTAGATATACGAAGC 121

Qy 1202 ttgcccgtcagctttccgcttttgcactcttatggttgcccttctcaccggtggcctgtggc 1261
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 120 TAAGTGGAAATGCTTGCATCAATTTTTCCTAATGCTCGGAATTTCTCATCGGAGTCCGCACTA 61

Qy 1262 ccggtgttattgagcaactctcgtagaagccaagtatctttataatattatagcatta 1320
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 60 CACCCATTGCTGCATGGCGGTGGACTCGATAGGAAGTAGAAGAACATGACTGTATAA 2
```

```

RESULT 14
CNS005TE          997 bp      DNA      linear      GSS 03-JUN-1999
LOCUS             Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION        BACR12K22 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION         AL060767
VERSION            GSS.
KEYWORDS           fruit fly.
SOURCE            Drosophila melanogaster
ORGANISM           Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydrioidea; Drosophilidae; Drosophila.
1 (bases 1 to 997)
Genoscope.
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT           Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osogawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1..997
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR12K22"
                     /note="end : TET3"
BASE COUNT          89 a 99 c 13 g 258 t 538 others
ORIGIN
661 acaagcaagattctaccacatcacacatacaaaagaatggaaatcgcgaaagcg 720
Db 356 ANNNNAANNAANNAANNAANNAANNNNTNANNNAANNAANNAANNTATGNNNN 415
Qy 721 gaacccagacagcgctccatccattctgttgaccacattcacaaactgttatggc 780
Db 416 NNNNNNNAANNAANNAANNAANGCTNNNNNNNNNNNNNNNNNNNAYYYTCTY 475
Qy 781 ctctcaatgttggtgtgtgtcggttactctcacaaatcttccctgttatgatgcc 840
Db 476 TTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTY 535
Qy 841 gttaccactgtgagattccggttctctaaacaaattatgtctgaaacagatgaatc 900
Db 536 TYCYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTY 595
Qy 901 tacatttgtcacaaagtctctcgcttcaattgttcgctgcgattggtatcatagtt 960
Db 596 CYTYCYTYCYTYCYTYCYTYCYTYCYTYCYTYTYTYTYTYTYTYTYTYTYTY 655
Qy 961 gctccaagattcactggcgacaccccggttacctcaaatgttcataatcttgctgct 1020

```

```

Db 656 YYAYKYCYYYYYTCTCYYYCYCYCYYYTCTCYCYCYYYTCTCTCTYTYTYTYT 715
Qy 1021 cttttcattcattctctctcttctgcaactatcggtccagacgcgcttactctgtt 1080
Db 716 YTCYCYYYYYTYTYTYTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 775
Qy 1081 ttctttgagtcactgacattttgttgatggtggaatgcacatgctttttcacatgga 1140
Db 776 YYCCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCY 835
Qy 1141 tacctcagcgctctggaatgggatacactccaaacgcgtgcacatctcactactcaaga 1200
Db 836 YCTYCCCTCTYTYTYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCY 895
Qy 1201 ttccctcagcttccggttcgcttgcactcttgccttgccttccacgcgtgctg 1260
Db 896 TYCYCYCYCCCTCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCYCY 955
Qy 1261 ccggttgattgagcacttcgt 1283
Db 956 TTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTYTY 978

RESULT 15
AA406898          258 bp      mRNA      linear      EST 01-MAY-1997
LOCUS             M8AFCZ7F07T3 Brugia malayi adult female cDNA (SAW96MLW-BMAF) Brugia
DEFINITION        malayi cDNA clone AFCZ7F07 5', mRNA sequence.
ACCESSION         AA406898
VERSION            AA406898.1 GI:2064981
KEYWORDS           EST.
SOURCE            Brugia malayi.
ORGANISM           Brugia malayi
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Brugia.
REFERENCE          1 (bases 1 to 258)
AUTHORS            Blaxter M.L., Waterfall, M., Daub, J., Lizotte, M., Baron, L. and Jones
S.J.
TITLE             Genes expressed in adult female Brugia malayi
JOURNAL            Unpublished (1996)
COMMENT           Contact: Blaxter ML
Institute of Cell, Animal and Population Biology
University of Edinburgh
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
3JT, UK.
Tel: +44 131 650 6760
Fax: +44 131 670 5450
Email: mark.blaxter@ed.ac.uk
The ABI trace of this sequence can be viewed at
http://www.sanger.ac.uk/brugia/AFC/M8AFCZ7F07T3.html
Seq primer: T3
FEATURES             Location/Qualifiers
     source          1..258
                     /organism="Brugia malayi"
                     /db_xref="taxon:6279"
                     /clone="AFCZ7F07"
                     /clone_lib="Brugia malayi adult female cDNA (SAW96MLW-BMAF)"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="XL1-Blue MRF/"
                     /note="Vector: Lambda Uni-ZAP XR; Site1: Eco RI; Site2:
Xho I; Lymphatic filarial nematode parasite of humans.
mRNA was prepared from approximately 50 adult females
isolated from the peritoneal cavity of birds and
converted to double-stranded cDNA using reverse
transcriptase and oligo(dT) followed by RNase H and DNA
pol I. The library has 5 x 105 independent recombinants
and the average insert size is ~900bp. The library was
constructed by Michelle Lizotte-Waniewski. The
library is available from Dr.S.A.Williams, email:
genome@smith.edu."
BASE COUNT          66 a 46 c 94 t

```

## ORIGIN

[illegible]

Search completed: July 23, 2002, 17:35:40  
Job time: 3563 sec

**THIS PAGE BLANK (USPTO)**

Result No.	Score	Query Match	Length	DB	ID	Description
1	100	100	100	100	100	100

Query Match 100.0%; Score 1388; DB 6; Length 1394;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1388; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	atgtaataatcaaccgacgaacacacattcccttgagcaggaagcatttcccaagagac	60
Db	1	ATGTAATAATCAACCGATCGAACACCTATGCCGTGAGCAGGAAGCATTTCCAAAGAGAC	60
Qy	61	aagtacaatatgtctactggctcgctcaattcttggattcggagttctctgccatgg	120
Db	61	AAGTACAATAATGTCTACTGCTCGCTCATTTCTGTGGATTCGGAGTCTTTCTGCCATGG	120
Qy	121	aatatgttcattactatcgccctcgagattattatgtgaattattggtcaaacccgagtcg	180
Db	121	AATATGTTCATTACTATCGCCCTGAGTATTATGTGAATATTGTTTCNAACCCGGATGGC	180
Qy	181	gtggagacatggatttcgaaagaattcatgggatctctttgacgattggctcacaacttcca	240
Db	181	GTGGAGACATGGTATTTCGAAAGAATTATGGGATCTTTGACGATTTGGCTCACAACTTCCA	240
Qy	241	aacgaagcattaatgttttcaacactgttcctcattattgctggtccctgatatcagc	300
Db	241	AACGAAGCATTAATGTTTTCAACCTGTTCCTCATTTATTTGCTGGTCCCTGATCTACCGC	300
Qy	301	gtcttggctcggtttgttcaacatcgtaacctgacaatcattctcatctcgtcatt	360
Db	301	GTCTTTGCTCGGTTTGCTTCAACATCGTCAACCTGACAATCATCTCATCTCGTCATT	360
Qy	361	gtctggagccactgaagattccatgctcctgggttttctgggtaacctcttgggaatggcg	420
Db	361	GTCTGGAGCCCACTGAAGATTCCATGCTCCCTGGTTTCTGGGTAACTCTTTGGAATGGCG	420
Qy	421	acttcaatcaatttagcaatgggctatatgaaacactcggtttatgagttggtggcgat	480
Db	421	ACTTCAATCAATTTTAGCAATGGGCTATATGAAACCTCGGTTTATGGAGTTGGTGGCGAT	480
Qy	481	ttccgcacacctacattggcgctctctgattggaacaacatttgcgatttctgata	540
Db	481	TTTCCGCACACCTACATTGGCGCTCTCTTGATTGGAACAACATTTGCGGATTCCTGATA	540
Qy	541	acggttggaaaaatcgagtgagcattttctggaatgatgagcctaaacttgggtgaatc	600
Db	541	ACGGTTGTGAAAATCGGAGTGAACCTATTTCTGGAATGATGAGCGCTAAACTTTGTGCAATC	600
Qy	601	gtctatttcggcatatggttggtgatcctctggtggtgcaaatgcaacttttcttattc	660
Db	601	GTCATTTCGGCATATCGTTGGTGATCCTTCTGTGTGTGCAATTCACATTTTCTTTATC	660
Qy	661	acaaagcaagatttctaccacatcacatcaaaaagaatggaattcgcgaaaaggcg	720
Db	661	ACAAAGCAAGATTTCTACCACTATCACATCAAAAAGGAATGGAAATTCGCGAAAAGCGG	720
Qy	721	gaaaccgacagacgctctccatccattctttggaccacattcacaactgttattgggcaa	780
Db	721	GAACCGACAGACGCTCTCCATCCATTTCTTGGACCACATTCACAAAAGTATTATGGCAA	780
Qy	781	ctcttcaatgttggttctgctgttgcggttactctcacaactctccctgttatgatgcc	840
Db	781	CTCTTCAATGTTGGTTCTGCTTTGGCGGTACTCTCAAAATTTCTCCCTGTATTATGACC	840
Qy	841	gttaccactgtagagattccggcttctcaaaaaaattatgctgaaacgatgaaatc	900
Db	841	GTTACCACCTCGTAGATTCGGGCTTCTTAACAAAATTTATGCTGAAAACGATGAATC	900
Qy	901	tacatttgcacaagtttctctgcttccaaatttctcgtcggtgattggatccatagtt	960
Db	901	TACACTTTGCTCAAAAGTTTCTCCTCGTCTTCAATTTGTTGCTGCGATTGGATCCATAGTT	960
Qy	961	gttcccaagattcactggcgcaaccccggttacctcaaatgttgcataatcttgcgtc	1020
Db	961	GCTTCCAAGATTCTACTGGCGACACCCCGTTACCTCAAAATTTGCCATAAATCTTCGCTGCT	1020

Qy	1021	ctttcattccattcttcttcttctgcaactatcggtgtccagacggtggttactctgtt	1080
Db	1021	CTTTTCATTCCATTCTCTCTCTTCTGCAACTATCGTGTCCAGACGGGTGCTTATCCTGTT	1080
Qy	1081	ttctttgagtcactgacacatttttggattggtggaattgccaatgtctttttccacatgga	1140
Db	1081	TTCTTTGAGTCTACTGACATTTTGTGATTGTTGGAATTGTCATGCTCTTTTTCACATGGA	1140
Qy	1141	tactcagcgtctcggcaatgggatcacactcccaaacgctcgtgccatctcactactcaaga	1200
Db	1141	TACTCTACGCGTCTGGCAATGGATACACTCCAACCGTGTGCCATCTCACTACTCAAGA	1200
Qy	1201	tttcggtcagcttccggtttgcactcttattggttggcctcttccacgggtggcctgctgg	1260
Db	1201	TTTCCGCGTCAAGCTTCGGTTGCACTTATGTTGGCTTCTCACGGTGGCTGTGG	1260
Qy	1261	ccggttattgagcacttcggtggacaagccaagtatcttataaattatttatagcatta	1320
Db	1261	CCCGTTGTTATGAGCACTTCGTGGACAAGCAAGTATCTTATAAATATTATAGCATTA	1320
Qy	1321	gagtatactgttatattgtttttttttttttaagctgtggaataaaaaattattaaaaaaa	1380
Db	1321	GAGTATACCTTGTATATGTTGTTTATTAAGCTGTGGAATAAAATATTATTAAAAAAA	1380
Qy	1381	aaaaaaaa 1388	
Db	1381	AAAAAAAA 1388	

RESULT 2

U55376/c 31049 bp DNA linear INV 09-AUG-2001  
LOCUS Caenorhabditis elegans cosmid F16H11, complete sequence.  
DEFINITION U55376  
ACCESSION U55376  
VERSION U55376.1 GI:1280130  
KEYWORDS HTG.  
SOURCE  
ORGANISM

REFERENCE

1 (bases 1 to 31049)  
The C. elegans Sequencing Consortium.  
Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium  
Science 282 (5396), 2012-2016 (1998)

JOURNAL

93069613  
2 (bases 1 to 31049)  
Wu, X.

REFERENCE

The sequence of C. elegans cosmid F16H11  
Unpublished  
3 (bases 1 to 31049)  
Waterston, R.

JOURNAL

Unpublished  
Direct Submission  
4 (bases 1 to 31049)  
Waterston, R.

JOURNAL

Submitted (18-APR-1996) Robert Waterston  
5 (bases 1 to 31049)  
Waterston, R.

REFERENCE

Submitted (09-AUG-2001) Department of Genetics, Washington  
University, Genome Sequencing Center, 4444 Forest Park Avenue, St.  
Louis, MO 63110, USA

JOURNAL

Submitted by:  
Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis, MO 63110, USA, and  
Sanger Centre, Hinxton Hall  
Cambridge CB10 1RQ, England  
email: rwenematode.wustl.edu and jes@sanger.ac.uk

COMMENT

NOTICE: This sequence may not be the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

#### NEIGHBORING COSMID INFORMATION

The 5' cosmid is C46C11, 200 bp overlap; 3' cosmid is K02G10, 2500 bp overlap. Actual start of this cosmid is at base position 1 of CELF16H11; actual end is at 10112 of CELK02G10.

#### NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder (P. Green and L. Hillier, ms in preparation).

#### FEATURES

Location/Qualifiers

- 1..31049 /organism="Caenorhabditis elegans"
- /strain="Bristol N2"
- /db\_xref="taxon:6239"
- /chromosome="X"
- /clone="F16H11"
- complement(1239..2839)
- /gene="F16H11.3"
- complement(join(1239..1368,1421..1617,1669..1878,1932..2131,2180..2466,2516..2649,2693..2839))
- /gene="F16H11.3"
- /note="weak similarity to HNP36 protein"
- /codon\_start=1
- /product="Hypothetical protein F16H11.3"
- /protein\_id="AAA98003.1"
- /db\_xref="GI:1280131"

gene

CDS

gene

CDS

gene

CDS

complement(join(10470..10481,10529..10624,10701..10861,10907..11025,11086..11396,11509..11613,12150..12431,13147..13430,13366..13631,13796..14006))

/gene="F16H11.5"

/note="contains a zinc finger of the C4-type; coded for by the following C. elegans cDNAs: yk347g3.3, yk347g3.5"

/codon\_start=1

/product="Hypothetical protein F16H11.5"

/protein\_id="AAA98005.1"

/db\_xref="GI:1280133"

/translation="MTHTFSINRTLFFSETTSKISMSCLVCETDAHQHGFIRCCRAC AAFRRRLTMNLRYKFRDKRCEVSNFKRYSRCRCRYEKCVRGVMRKDNLTSPVSVVQ NKSETEKDSQNETDSIGHSPHSSLESTCRPPSEFOQOQNOQNDPPLLMMNGNTEFVO EVASDAYQPSNDMHQNLNFSFRTVIEVGRAVMTPTFPQPPPIHHPQDFDYQTDLLS TDQNSAIPSSFTQYNQVPOEYNNESSGNQMFAMAQAQATQDLHVNVQVLTPTID AMFQGPDPACQLEDIPLTCQKQALLAYREHNKQMPDQDKMIENPLDMMENFRNHYI EIEHIAFCMSIRVFAQLPKDQKWIFKHEWTRFYEIDRCFATQCRQNGYNTDREGLT LMGQIINFGISVNLKLENTSDMATQVRNFKGSKMDKRLIINFPKKLQPTFYEELMYM MMSMTWSVSNLPGITDRTDRISEKVELRAEDLHTYYAEQYDNNNNPNYAGRTIRLSI ASAYDEITERKREDSQVSTNFIKNDFFSDLTDFPCLN"

/gene="F16H11.2"

16440..16613

16440..16613

/gene="F16H11.2"

/codon\_start=1

/product="Hypothetical protein F16H11.2"

/protein\_id="AAA98006.1"

/db\_xref="GI:1280134"

/translation="MTQDRTEERERERERERERERERERERERERERERERE RERERESNRKR"

21734..26707

/gene="F16H11.1"

join(21734..21787,21843..21909,21958..22079,24004..24109,24156..24295,24739..24809,24855..24966,25036..25127,25173..25257,25305..25397,25441..25536,25584..25751,26297..26394,26451..26593,26649..26707)

/gene="F16H11.1"

/note="similar to melibiose carrier protein (thiomethylgalactoside permease II); coded for by the following C. elegans cDNAs: cm01e2, cm21e6"

/codon\_start=1

/product="Hypothetical protein F16H11.1"

/protein\_id="AAA98007.1"

/db\_xref="GI:1280135"

/translation="MTDEETGDPVVRSCDDIVIPRKSIRPDHDIISTKKIARGIGH FYNDLCASMMFTYFMIYNEKVLKFOSSRAGMLIGQVTDISTPLVGIFSDSNILPA CFQJIGRMSHLIGTIVLSLSPFMIENKCFKSTSEWLKVLWFVFIMVFOFGWA SVQIISHLALIPELSSPASRATMNSLRVAVIANLSYFALWLLSESTGHTSIGPW DFSHERAGLVVVLGITVAVFEVAFREPTNYRFRSLSSDSDSLVRMHTSWF GHVQFQIALLMLSLRYLINISQVFFPYITMONTQEKYVAILPHVAYLSSFSVSW NSLPVSKSKILISFGLASMLSCFMMDLDPGWKIYALAVGIIAQAILLITSLS ITADLINNKTESGAFVYGMSFFDKLSNGIAYQLIETWLPAYDALKPHEVAIFRYR MVFVPGCLVLAFILVLLSLAPFKIGERRRARRPEQAIMEQDDIYPEIV"

BASE COUNT 10124 a 5100 c 5591 g 10234 t

ORIGIN

Query Match 72.9%; Score 1012; DB 3; Length 31049;  
Best Local Similarity 82.0%; Pred. No. 1.9e-259;  
Matches 1371; Conservative 0; Mismatches 5; Indels 296; Gaps 6;

Qy 1 atggtataatacaaccgatcaaacacctatgccgttgagcagaagcatttccaagac 60  
|||||  
Db 2839 ATGGTAATAATCAACCGATCGAACACCTATGCCGTGAGCAGGAGCATTTCCAAGAGAC 2780  
|||||  
Qy 61 aagtaacaattgtctactgctcgtcatttctgtgattcggaggtctcttcgcatg 120  
|||||  
Db 2779 AAGTACATATGCTACTGGCTCGTCATCTTGTGGATTCGGAGTCTTCTGCCATGG 2720  
|||||  
Qy 121 aatagtcttactatcatcgccctg----- 145  
|||||  
Db 2719 AATAGTTCAATTACTATCGCCCTGAGGTTTGCAACTAGATAATTTTCACTAATAGAGTT 2660  
|||||  
Qy 146 -----agttattatggaattattgttcaaacccgatggcgtggagacattgtattc 197  
|||||  
Db 2659 CCGTGTTCAGTATTATGTGAATTATTGTTCAACCGGATGCGGTGGAGACATGTTATTC 2600  
|||||  
Qy 198 gaaagaattcggatcttgacgatggctcacaacttccaaacgaagcattaaagt 257  
|||||  
Db 2599 GAAAGAAATCAGGGATCTTGACCATGCCCTCACAACTTCCAAACGCAAGCATTAAATGT 2540  
|||||  
Qy 258 ttcaacctgttctcattattgc----- 281

Db 2539 TTTCAACCTGTTCCCTCATTAATGCGTAAGTTACAGTTCTCTTAAACTCTCAAAAATCAAT 2480  
QY 282 -----tggtccctgatctaccggtgttctgctcgggttcttgcattcgaatcg 328  
Db 2479 CAATATTTTTTATGAGTGGTCCCTTGATCTACCGGGTCTTTGCTCCGGTTGCTTCAACATCG 2420  
QY 329 tcaacctgacaatctctcattccctcgctcattgtctgagccacctgaagattccatgt 388  
Db 2419 TCAACCTGACATCAATTCATCTCGTCGTCATGTGTTCTGGAGCCCACTGAAGATTCATGT 2360  
QY 389 cctggttttcttggtgtaacctcttggaatggcgaccttcaatcaatttttagcaatggctat 448  
Db 2359 CCGTGTCTTTCTGGGTAACTCTTGAATGGCGACTTCAATCAATTTTAGCAATGGGCTAT 2300  
QY 449 atgaaaaactcgggttatgaggtggtggtggttttccgcacacctacattggcgctctct 508  
Db 2299 ATGAAAACTCGGTTTATGGAGTTTTCGGGATTTTCCGACATTTTCCGACACCTTACATTTGGCGCTCTCT 2240  
QY 509 tgattggaaaacaacatttgcggtatctgataaacggttctgaaaaatcggagtgacctatt 568  
Db 2239 TGATTTGGAACACAAATTTGCGGATTTGCTGATAACGGTTGTGAAAATCGGAGTGACCTATT 2180  
QY 569 -----ttctggaatgatg 580  
Db 2179 GTAAGTAATTTTCAATTAATTTGCTTTAAAAACTATAACATTTTATGTTCTGTAATGATG 2120  
QY 581 agcctaaactgttgcacatgctctatttctgcgcatactgttggatccttctggtgtg 640  
Db 2119 AGCCTAAACTTGTTCGAATCGTCTATTTTCGGGATATCGTTGGTGATCCTCTCGGTGTG 2060  
QY 641 caattgcactttcttcatcaaaagcaagatttctacacatcaccatcaaaaaggaa 700  
Db 2059 CAATTTGCATTTTCTTTATCAAAAGCAAGATTTCTACCACTATCAACATCAAAAAGGAA 2000  
QY 701 tggaaattcggaaaaagcggaacccgacagacctctccatcttcttggaccacat 760  
Db 1999 TGGAAATTCGGAAAAAGCGGAACCCGACAGACCGCTCTCCATCTTCTTGGACCAT 1940  
QY 761 tcacaac----- 768  
Db 1939 TCACAAACGTGNGTTTTTAAATCAATTTAGCTTCAAAATAAATATGTGTGACTAATTTCA 1880  
QY 769 -tggtatgggcaactcttcaattgttgggtctgtctggtccgttactctcaaatctccc 827  
Db 1879 GTGTTATGGCAACCTCTCAATGTTGTTGTTCTGCTTTGCGGTTACTCTCACAATCTTCCC 1820  
QY 828 tgttatgatgaccttacacctcgtggtggtatccggtctcctaaacaaaattatgtctga 887  
Db 1819 TGTATGATGACCGTTACCACTCGTGGAGATTCCGGCTTCCCTAAACAAAAATTATGCTGA 1760  
QY 888 aaacgatgaatctacactttgctcacaaagtttctcgtcttcttcaattgttgcgtgcgat 947  
Db 1759 AAACGATGAATCTACATTTGCTCACAAAGTTTCTCGTCTTCAATTTGTTGCGCTGCGAT 1700  
QY 948 tggatccatagttgttccaaagattcactg----- 977  
Db 1699 TGGATCCATAGTTGTTCCAAAGTTTCACTGGGTAAATTTGAAATTCAGAAATAACATTGCT 1640  
QY 978 -----gcgacaccocggttacctcgaatttgcataatcttgcg 1016  
Db 1639 TGAATPATCATCAAAATTTGCGAGCCGACACCCCGTTTACCTCAAAATTTGCCAATCTTCCG 1580  
QY 1017 tgccttttcttcttctctctctcgaactatcgttccagacgctgcttctcc 1076  
Db 1579 TGCCTCTTTTCATTCATTTCTCTCTCTGCAACATATCGTGTCCAGACCGCGTGTATCC 1520  
QY 1077 tgtttcttcttgagtactgacatttttctgattggtggaattgccaattgtctttttcaca 1136  
Db 1519 TGTTTCTTTTGTAGTCTACTGACATTTTGTGATTTGGTGAATTTGCCANGTCTTTTTCACA 1460  
QY 1137 tggatacctcagcgtctctggaatgggatacactcaaa----- 1175  
|||||

Db 1459 TGGATACCTCAGCGCTCTGGCAATGGGATACACTCCAAAGTGAGTTACAATTCGGTTGTT 1400  
QY 1176 -----cgctgctccatctcactactcaagatttg 1204  
Db 1399 ACTATGAATTTGGAATCTTTAAATTTTATAGCGTGTGGCATCTCTACTACTCAAGATTG 1340  
QY 1205 ccgctcagctttcccttctgacctcttattggttggcctctcaccggtggtggtggtgccc 1264  
Db 1339 CGGCTCAGCTTTCCTGTTGCACTCTTATGTTGGCTTCTCACCCTGTCGCTGTGGCG 1280  
QY 1265 ttgttattgagcaactcgtggacaagccaagtattcttataataattatttagcattagat 1324  
Db 1279 TTGTTATTGAGCAGCTTCGTGGACAAGCAAGTATCTTATAAATATTATAGCATTAGACT 1220  
QY 1325 atactgttatgttgttttttatttaactgtggaataaataattattaaa 1376  
Db 1219 ATACTGTATATGTTGTTTATTAACTGTGGAATAAAATAATTATTAAA 1168  
  
RESULT 3  
AY058347 AY058347 2065 bp mRNA linear INV 16-OCT-2001  
DEFINITION Drosophila melanogaster GH12067 full length cDNA.  
ACCESSION AY058347  
VERSION AY058347.1 GI:16182780  
KEYWORDS FLY CDNA.  
SOURCE fruit fly.  
ORGANISM Drosophila melanogaster  
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
1 (bases 1 to 2065)  
Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,  
Champe,M., Chavez,C., Dorsett,V., Parfan,D., Frise,E., George,R.,  
Gonzalez,M., Guarin,H., Li,P., Liao,G., Miranda,A., Mungall,C.J.,  
Nunoo,J., Pacleb,J., Paragas,V., Park,S., Phouanavong,S., Wan,K.,  
Yu,C., Lewis,S.E., Rubin,G.M. and Celniker,S.  
Direct Submission  
Submitted (08-OCT-2001) Berkeley Drosophila Genome Project,  
Lawrence Berkeley National Laboratory, One Cyclotron Road,  
Berkeley, CA 94720, USA  
Sequence submitted by:  
Berkeley Drosophila Genome Project  
Lawrence Berkeley National Laboratory  
Berkeley, CA 94720  
This clone was sequenced as part of a high-throughput process to  
sequence clones from Drosophila Gene Collection 1 (Rubin et al.,  
Science 2000). The sequence has been subjected to integrity checks  
for sequence accuracy, presence of a polyA tail and contiguity  
within 100 kb in the genome. Thus we believe the sequence to  
reflect accurately this particular cDNA clone. However, there are  
artifacts associated with the generation of cDNA clones that may  
have not been detected in our initial analyses such as internal  
priming, priming from contaminating genomic DNA, retained introns  
due to reverse transcription of unspliced precursor RNAs, and  
reverse transcriptase errors that result in single base changes.  
For further information about this sequence, please visit our location  
and relationship to other sequences, including its location  
(http://fruitfly.berkeley.edu) or send email to  
cdna@fruitfly.berkeley.edu.  
Location/Qualifiers  
1. .2065  
/organism="Drosophila melanogaster"  
/strain="y; cn bw sp"  
/db\_xref="taxon:7227"  
/map="26E2-26E2"  
1. .2065  
/gene="BEST:LD04971"  
/note="alignment with genomic scaffold AB003614"  
/db\_xref="FLYBASE:FBgn0026585"  
347..1723  
/gene="BEST:LD04971"  
/note="Longest ORF"



[illegible][illegible]





```

source      1..174376
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/chromosome="2L"
/map="26F-27X"
/clone="BACR5K24 (D1294)"
/clone_lib="RPCr-98 (Roswell Park Cancer Institute
Drosophila melanogaster BAC library, partial EcoRI in
pBACE3.6)"
BASE COUNT 48681 a 38138 c 38499 g 49058 t
ORIGIN
Query Match      6.5%; Score 90.8; DB 3; Length 174376;
Best Local Similarity 45.4%; pred. No. 4e-13;
Matches 556; Conservative 0; Mismatches 632; Indels 36; Gaps 5;

QY 47 cattccaagagacaagtacaataattgtctactggtcgctcattctgttgatcgag 106
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 39300 CAGCGCAAAAGATAAATTCCTTATCGTGTCTTTATATCTCTGTCACGCGTGGCA 39359

QY 107 ttcttgcgatgaatatgttcaattactatgcgc-----tgagtattatgtaatt 160
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 39360 CGCTGATGCCATGAACATGTCATACGCGCAAGTCTATTTCGAGACATCAAGTTCG 39419

QY 161 attggttcaaacggatggcgtggagacatggtattcgaagaattcatggatcttga 220
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 39420 GACCAACACACCGTGGCCACGGAAGTAGTTACCGCACCCATTTCATGCAAAATATGG 39479

QY 221 cgattggctcacacttccaaacgaagcataatgttttcaacctgttccctcattattg 280
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 39480 GCATTGCCTCGAGATCCCTAAACCTGGTCTTCAACTGGCTGAACATCTTTGTCAACTTGG 39539

QY 281 ctggtcccttgatcacccggtcttctgcctcgggtttcgttccaaacatgcgaacctgacaa 340
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 39540 CGCGCGACCTGACACCCGGAATCGCTACAGCATCATCTTCGAAATGGTCATTCTGCTGG 39599

QY 341 tcattctcctctcgtcatgttctggagccacctgaagattccatgctcctgggtttct 400
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 39600 TTACCATATTGTGGCCATGCTCG-----ACTCTCGCAGTGGCGGGCGTCTTCTCT 39653

QY 401 gggtaactcttggaaatggcgacttcaatcatattgtaaatggcgtctatgaaactcgg 460
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 39654 GGACCACCATGGTGCATTTGTTGTTAAACGTGTGCAACGGCATCTACCAGAACACCA 39713

QY 461 ttatgagattggtggcgaatttccgcacacctacattggcgctctcttgattggaaca 520
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 39714 TATACGGAATAGTGGGCATCGTTGCCCATCAAGTACACAGCGCGCTGCTGTGGGCTCCA 39773

QY 521 acattcgcgatgctgataacggttgtgaaacLcgagigacatttttcgaaatgatg 580
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 39774 ACATCAGCGCTGTTCCACCACCGCAATGGCCTTGATATGGGGTGAGATCTTCTCGTCCA 39833

QY 581 agcctaaactgttgcaactcttcaatttcggcatatgctgttggtgatcctcttggtgtg 640
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 39834 AG---AGGACTTCGGCCATCTACTACTTCGTGACCGCCATCCCTAGTGTGCTGCTCTGCT 39890

QY 641 caattgcacttttttatcacaagaagattcttaccactatcacaactcaaaaaggaa 700
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 39891 TCGACACGTAATTTGGCTGCCGTGAACAAAGTTCTTCCGCGCATACGAGACCATTAGCC 39950

QY 701 tggaaattcgaaaaggcgaaacgcagacagcgtctccatctcttggaccacat 760
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 39951 GGAGCAGCGAANAAGAAATCTGATTTTAAGGCCACAGCTGACGTCGCCCTACTTGGCAGATCT 40010

QY 761 tcacaaactgttatgggcaactcttcaatttgggttctgtctgttgcgttactctcaaa 820
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 40011 TCAAGAAAGCAGCGCCCACTATTCAACATCTTCCTCAGCTTTTTCGTACGCTGTGTCGG 40070

QY 821 tcttccctgttatgatgacctttacaaactcgtggagattccggcttccctcaaaaaatta 880
|| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 40071 TTTTCCCGGGGATTCAGTCAAGAGTGCACCGCTCCGACCCAAATTTTCGTA----- 40120

```

Qy	861	tgtctgaacagatgaaattacacttctgtcccaagaattttctctcgttcttctaatttgctg	940
Dd	40121	---GTAGGCCCGGACTACTTTTACATGGTGCTTCTTCCGCAGCACTTTAACGTCCTCG	40175
Qy	941	ctgcgattgatcatcattagtcttcgaagattcactcgccagacacccccgttacctcaaat	1000
Dd	40176	CAATGCTGGGCGAGTTTAACCAATCATGTTGGGTGAATGCGCGGTTCCTCCTGCGG	40235
Qy	1001	ttgcacataattcttcgctgtctttttcattccattctcttcttcttcgcacaata-----tc	1054
Dd	40236	TGCCAGTGGTCTTCGCTGCGGTTCATCCCTTGTGTCATGTCGAACATACGTTCCGC	40295
Qy	1055	gtgtccgaacgctgtcttatctcgttttttttggagtcctactgacattttgtgattgggtg	1114
Dd	40296	CGGACTCGGTGCGTCTAATTGGCCGTATTCATCGAGAACAATGTTGGGTGTACTGGGGCATCG	40355
Qy	1115	qaattgccatgctctttttcacatgataacctcagcgtctctggcaatgggatacatactccaa	1174
Dd	40356	GCATCGCAATGGCTACAGCTCCGGGTATTCACCTCTCTGGGCATATGATGTACC GCCCGC	40415
Qy	1175	acgtctgccatctcactactcaagattttcgcgtcagctttccgtttgtcactcttattg	1234
Dd	40416	AGACGGTGCACCAAATACGACCACTGCCGAATGTACGCAGCTGCCATGCTGATCA	40475
Qy	1235	ttggcctcttcacagggtggcctgt	1258
Dd	40476	CGGGCATCTTCTCCGGAGTGTGT	40499

RESULT 7

AE003614	259464 bp	DNA	linear	INV 04-OCT-2000
LOCUS	Drosophila melanogaster genomic scaffold 142000013386055	section 7		
DEFINITION	of 63, complete sequence.			
ACCESSION	AE003614 AE002690			
VERSION	AE003614.2 GI:10728613			
KEYWORDS	HTG.			
SOURCE	fruit fly.			
ORGANISM	Drosophila melanogaster			
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
AUTHORS	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;			
	Muscomorpha; Ephyarodea; Drosophilidae; Drosophila.			
	1 (bases 1 to 259464)			
	Adams,M.D., Celisner,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,			
	Ananides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F.,			
	George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,			
	Sutton,G.G., Wortman,J.R., Vandell,M.D., Zhang,Q., Chen,L.X.,			
	Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,			
	Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor			
	Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,			
	Andrews-Pfannkuch,C., Baldwin,D., Ballew,R.M., Basu,A.,			
	Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,			
	Benos,P.V., Bertram,P.P., Bhandari,D., Bolshakov,S., Borokova,D.,			
	Botchan,M.R., Bouck,J., Brokstein,P., Bottier,P., Burtis,K.C.,			
	Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,			
	Cherry,J.M., Cawley,S., Dahke,C., Davenport,L.B., Davies,P., de			
	Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,			
	Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,			
	Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,			
	Fleischmann,W., Fosler,C., Gabriellian,A.E., Garg,N.S.,			
	Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,			
	Guan,P., Harris,M., Harris,N.L., Harvey,D., Helman,T.J.,			
	Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,			
	Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,			
	Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,			
	Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,			
	Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,			
	McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.Y., Mobarry,C.,			
	Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,			
	Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,			
	Nusskern,D., Pacle,J.M., Palazzolo,M., Piftomanoglou,S., Pan,S.,			
	Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,			

Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I., Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C., Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R., Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A., Weinstock,G.M., Weissbach,J., Williams,S.M., Woodage,T., Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F., Zaveri,J.S., Zhang,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H., Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O., Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.  
The genome sequence of *Drosophila melanogaster*  
Science 287 (5461), 2185-2195 (2000)  
20196006  
2 (bases 1 to 259464)  
Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.  
Direct Submission  
Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA  
On Oct 9, 2000 this sequence version replaced gi:7297129.  
Location/Qualifiers  
1. .259464  
/organism="Drosophila melanogaster"  
/db\_xref="taxon:7227"  
/chromosome="2L"  
join(538..1127,1190..1366,1421..>1745)  
/gene="CG9550"  
/product="CT27016"  
/db\_xref="FLYBASE:FBan0009550"  
/db\_xref="FLYBASE:FBgn0031826"  
<538..>1745  
/gene="CG9550"  
/map="26D7-26D9"  
/db\_xref="FLYBASE:FBan0009550"  
/db\_xref="FLYBASE:FBgn0031826"  
join(538..1127,1190..1366,1421..1745)  
/gene="CG9550"  
/note="CG9550 gene product"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0009550"  
/db\_xref="FLYBASE:FBgn0031826"  
/protein\_id="AAF52397.1"  
/db\_xref="GI:7297130"  
/translation="MWSLSRTSLTVACVGVYTYVLFIFLLPMLPDPVTLQSLV  
SYSTAEHRDLNLTLDAGQPIRSMLTVFGSGALTLDLNLAHQPGCYHYAPLIAYE  
SRSTAKEHRLDEALADYNDYNSAEIMQWSPVRFYGAQAKICRSYSQEA  
CWDPETMAACKLYPFINNAVYLNRLNLASLLERKDLNLLMLVDRDPRTYSRNN  
NKWCSDRDCEAOELCGDMVSDYQVETLTQAYPQRFSTIIRYEDLFTQPDSEIKQVFD  
FYGLPLERNTRIQKHPRSSYFQSSRMKELSSQSAFEWISMMVDDIGAVDVCQ  
AMDLGWRSIQDNRHFSFESQFIMDKAY"  
2069..>33227  
/gene="CG9551"  
/product="CT27018"  
/db\_xref="FLYBASE:FBan0009551"  
/db\_xref="FLYBASE:FBgn0031827"  
/evidence=not\_experimental  
<2069..>33227  
/gene="CG9551"  
/map="26D7-26D9"  
/db\_xref="FLYBASE:FBan0009551"  
/db\_xref="FLYBASE:FBgn0031827"  
/evidence=not\_experimental  
2280..3327  
/gene="CG9551"  
/note="CG9551 gene product"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0009551"  
/db\_xref="FLYBASE:FBgn0031827"  
/evidence=not\_experimental  
/protein\_id="AAF52398.1"  
/db\_xref="GI:7297131"  
/translation="MSTGICSSSSTEEQOHLLEADLACSLLGPODPSNVDTGTRR  
AKORLSNGKRERRSRNRANLIGICVSSICILLIATTTQHRPLTTTFNQSGAGROG  
ASVDGFSNLSANSAFYDRQQOQQQSSQATVYNLTATIDVLSAQSRILAEME  
NFEPYRGAERLDTMTPTETGTPVRSVVVTSWRSGSTFLGLDILNSIPNGFYHEPLLD

mRNA

gene

CDS

mRNA

gene

CDS

mRNA

gene

FGIKQIRDPDDQLAVQNLKLLNCDYADMIDYLNFGKTHTYLFEHHTRLWDVCREFP  
RFCWRPAFLTRFCRLFTIQSMKTVRLRLAQAEKLEQOR"  
join(<27409..27443,28126..28308,28377..28578,  
28641..>28772)  
/gene="CG13765"  
/product="CT33249"  
/db\_xref="FLYBASE:FBan0013765"  
/db\_xref="FLYBASE:FBgn0031828"  
/evidence=not\_experimental  
<27409..>28772  
/gene="CG13765"  
/map="26D9-26E1"  
/db\_xref="FLYBASE:FBan0013765"  
/db\_xref="FLYBASE:FBgn0031828"  
/evidence=not\_experimental  
join(27409..27443,28126..28308,28377..28578,28641..28772)  
/gene="CG13765"  
/note="CG13765 gene product"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0013765"  
/db\_xref="FLYBASE:FBgn0031828"  
/evidence=not\_experimental  
/protein\_id="AAF52399.1"  
/db\_xref="GI:7297132"  
/translation="MVMPQHVVISPIILSSRVIVLLVRDPRTMQSRHRVWCGGNEDC  
EDPRLVCQDLRDDYKTAEVLLKLYPSRFVRYEDLSLSPSEMTQDILFYGLPDPDA  
VEEFLDTHTKVNIIGVSVSTRDSRAPFHMQLKPEIKIQDVCTEAMDLGWYRI  
ENFNFTTQOTDPMVMPPT"  
complement(join(30215..30949,31026..31270,31333..33087,  
35664..35710,38587..38869))  
/gene="cli"  
/note="Nucleotide sequence of the Celera sequence differs  
from the published sequence for this transcript."  
/product="CT27024"  
/db\_xref="FLYBASE:FBan0009554"  
/db\_xref="FLYBASE:FBgn0000320"  
complement(<30215..>38869)  
/gene="cli"  
/note="CG9554"  
/map="26D9-26E1"  
/db\_xref="FLYBASE:FBan0009554"  
/db\_xref="FLYBASE:FBgn0000320"  
complement(join(30768..30949,31026..31270,31333..33087,  
35664..35710,38587..38658))  
/gene="cli"  
/note="cli gene product; Nucleotide sequence of the Celera  
sequence differs from the published sequence for this  
transcript"  
/codon\_start=1  
/db\_xref="FLYBASE:FBan0009554"  
/db\_xref="FLYBASE:FBgn0000320"  
/protein\_id="AAF52400.1"  
/db\_xref="GI:7297133"  
/translation="MVTILMPYNYAAPRCGLIDKMIPEKVKRPKTDHTDTHBRNLCNL  
SQQQQQQQQQOQTHQQQQQQQQSHSQSHSTVLASNGPSSAGAGVGVGGGGSG  
GGVGGVGCQSPGLPQSQPLPTIGSLASLGHYSGNANPNVNSSCSLATASFF  
AQSAGSSFTYQOAGGTSGVSGRDVGGAATVMSHTHDGTGSAVKSRSRSPGOV  
HASLDGSAVGSNLYGCSSASNPLDGAVAVNSAAVAAAAYVDGKHDIYYVNSMQ  
YTPPPFYSGYGTYYAATAARQAQMEPCAAAAYLTPSYAASGNNSQLYSPYAG  
YTNFGQDYGITYNEQIGNYSPANYSPYAVSSPSSASAHGHGFHVAASSNLSEPTD  
YNTPTVQHTHSPSPSPISPTSGIGPLGNVSAASAAAALNSGGSSVGTAGSGG  
VATSKTPTGKTGRGRRHQQPSPTSTADTGNSEAVKPERFVWDLDELIIIFH  
TLISGSYANRYTKDHSILMTAFRMEEMVFNADTHFFNEIECDOVHIDVSDSDN  
GODLSAYNFATDGFHTPPGAPNLCILPTGVRGVDMRMKLAFRYRKIKDIYSYRG  
NVGTLGLPGKREAWLQIRSEIEVATDNWATLALACLMSIQRENCVNVLVSTQLAPA  
LAKVLLGLGIGFNIENIYSAKHGHCYERIVTRFGKSTYVYVIGDNGEEETAAGA  
MNEFPWRISAHSDIRALYTALDMGFL"  
complement(join(<58570..59115,59172..59817,60078..>60118))  
/gene="CG9595"  
/product="CT27132"  
/db\_xref="FLYBASE:FBan0009595"  
/db\_xref="FLYBASE:FBgn0031829"  
complement(<58570..>60118)

```
/gene="CG9595"
/map="26E1-26E2"
/db_xref="FLYBASE:FBgn0009595"
/db_xref="FLYBASE:FBgn0031829"
complement(join(58570..59115,59172..59817,60078..60118))
/gene="CG9595"
/contig="CG9595"
/codon_start=1
/db_xref="FLYBASE:FBgn0009595"
/db_xref="FLYBASE:FBgn0031829"
/protein_id="AAF52401.1"
/db_xref="GI:7297134"
/translation="MMHRKLVNNVNEQDAELRMRILARKVIFVLAGPODRFTDEF

Query Match      6.5%; Score 90.8; DB 3: Length 259464;
Best Local Similarity 45.4%; Pred. No. 4.1e-13;
Matches 556; Conservative 0; Mismatches 632; Indels 36; Gaps 5;

Qy 47 catttccaagagacaagtacaaatattgttctactgctgctgctcattcttgttgattcggag 106
Db 65601 CAGCGCCAAAGATAAATCCCTTATCGTGTCTTTATATTCCTGCTGCACGGCGTGGCA 65660

Qy 107 ttctctccaatgaatagtctcattactatcgcccc-----tgagtattatgtgaatt 160
Db 65661 CGCTGATGCCATGGAAACATGTTTCATCACGGCCAAAGTCCCTATTTCGAGGACTTCAAGTTCCG 65720

Qy 161 atttggttcaaacggatggcggtgagacatggtatttcgaagaattcatggatcttga 220
Db 65721 GACCAAAACACACCGTGGCCAGGGAAGTGAGTTACCGACCCCATTTTCATGCAAAATATGG 65780

Qy 221 cgattggttcacaacttccaaacgcaagcattaatgttttcaacctgttcctcattattg 280
Db 65781 GCTTTGCTCGCAGATTCTTAACCGTGTCTTCAACTGGCTGAACATCTTTGTCAACTTTG 65840

Qy 281 ctggtccctcgtatctacgcgtcttctgctcgggttcttgaacatcgcacacctgacaa 340
Db 65841 CGCGGCACTGACCAACCGAATCGTCTACAGCATCATCTCGAAATGTCTATCTTCGTCGG 65900

Qy 341 tcattctcatctgctcattgtctcgtgagccacgtgaagattccatgctcctgggtttct 400
Db 65901 TTACCAATATTTTGGCCATGCTCG-----ACTCTCCAGTGGCGGGCGTCTTCTTCT 65954

Qy 401 gggtaactcttggaatggcgacttcaatcaatttttagcaatgggctatatgaaactcgg 460
Db 65955 GGACCACCATGCTGNGCATTTGTTGTTAAACGCTGTGCAACGGCATCTACAGAACACCA 66014

Qy 461 tttatggagttggtggcgatttccgcacacccatcattggcgctctcttattgtagaaca 520
Db 66015 TATACGGAATAGTGGCATCTGTTGCCCATCAAGTACACAGCGCGCTGCTCTGGGCTCCA 66074

Qy 521 acatttgcggtattgctgataacgggttgtaaaatcgagtgacctattttctgaatgatg 580
Db 66075 ACATCAGCGGTGCTTCAACACCGGCATGGCTTGATATGGGGTGAGATCTTCTCGTCCA 66134

Qy 581 agcctaaacttgtgcaatctctatttggcgcataatcgttggtagctccttgggtgtg 640
Db 66135 AG---AGGACTTCGGCCATCTACTTCTCGTACCGCCATCCTAGTGTGCTGCTCTGCT 66191

Qy 641 caattgcaatttttattatcaaaagaagatttttaccatatacaccatacaaaaagaa 700
Db 66192 TCGACACGATCTATTCGCTGCGCTGCGTGAACAAAGTTCTTCGGCACTACGAGACCATTAGCC 66251

Qy 701 tggaaattcgcgaaaaggcgaacacgacagacgctctccatccattctttggaccacat 760
Db 66252 GGAGCAGCGAAAGAAATCTGATCTAAGGCCACGAGTGAAGTGCCTTACTTGGCAGATCT 66311

Qy 761 tcacaaactgttattgggcaactcttcaatgtttggttctgtgttggccttactctcaaa 820
Db 66312 TCAAGAAAGCAGCGCCCAACTATTCAACATCTTCCCTCACGCTTTTTCGTACGTTGTCGG 66371

Qy 821 tcttccctgttatgatgaccgttaccactcgtggagattcggcttccctaaacaaatta 880
Db 66372 TTTTCCCGGCGATTTCAGTCGAACGTGCACCGTCCGACCCCAAAATTCGTA----- 66421
```

```
Qy 881 tgtctgaaaacgatgaaatctacactttgtctcacaagtttctcgtcttctcaattgttcg 940
Db 66422 -----GTAGGCCCGGACTACTTACACTGGTCACTTGTCTCGCGACCTTTTACGTCCTCG 66476

Qy 941 ctgogattggatccatagttgcttccaagattcactgagccgacacccogttaccctcaaat 1000
Db 66477 CAATGCTGGCAGATTTAAACACATCGTGGTGCATATGGCGGGTCCGAGGTTCTCTGGG 66536

Qy 1001 ttgccaataattgogtctcttttccattccaattcttcttcttcttctgaacta-----tc 1054
Db 66537 TGCCAGTGGTCTTGGCTGTGGCTTCATCCCTCTGTTTGTCTATGCACTACGTTCCTCGC 66596

Qy 1055 gtgtccagacgcgtcttctatctctgtttttttgagttctactgacatttttggattggg 1114
Db 66597 CGGACTCGTGGCTCATCTGGCCGTATTTCATCGAAGCATGGGTGTACTGGGCGCATCG 66656

Qy 1115 gaattgccatgtctttttccatgataccctcagcgtcttgcaatggatagatacactccaa 1174
Db 66657 GCATGCAATGGCTTACAGCTCCGGGTATCTCAGCTCTCTGGGCATGATGTACGCCCGCG 66716

Qy 1175 acgtgtgcactctcactactcaagatttgcgctcagcttcttccgttttgcactcttatgg 1234
Db 66717 AGACGGTGCACACCAATACCAGACCACCTGCGGGAATGTACGACGTGCCATGCTGATCA 66776

Qy 1235 ttggcctctcaccggtggcctgt 1258
Db 66777 CGGGCATCTTCTCCGAGTGTGT 66800

RESULT 8
CEFA4D12 CEF44D12 31853 bp DNA linear INV 11-DEC-2001
LOCUS Caenorhabditis elegans cosmid F44D12, complete sequence.
DEFINITION Z68298
ACCESSION Z68298
VERSION Z68298.1 GI:1130621
KEYWORDS HTG; Serine/threonine protein kinase; Tight junction protein; Zinc finger protein.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE 1 (bases 1 to 31853)
AUTHORS none.
TITLE Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium
JOURNAL Science 282 (5396), 2012-2018 (1998)
MEDLINE 99069613
REMARK The C.elegans Sequencing Consortium.
REFERENCE 2 (bases 1 to 31853)
AUTHORS Coles, L.
TITLE Direct Submission
JOURNAL Submitted (20-DEC-1995) Nematode Sequencing Project, Sanger Institute, Hinxton, Cambridge CB10 1SA, England and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or rwenematode.wustl.edu
COMMENT Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other available information.
Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is not the entire insert of clone F44D12. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
The true left end of clone F44D12 is at 1 in this sequence. The true right end of clone F44D12 is at 6268 in this sequence.
The true left end of clone T04B2 is at 31750 in this sequence. The start of this sequence (1..104) overlaps with the end of sequence
```

268882.  
The end of this sequence (31750..31853) overlaps with the start of  
sequence Z68299.  
For a graphical representation of this sequence and its analysis  
see: - [http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?](http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=F44D12)  
name=F44D12  
IMPORTANT: This sequence is NOT necessarily the entire insert of  
the specified clone. It may be shorter because we only sequence  
overlapping sections once, or longer because we arrange for a small  
overlap between neighbouring submissions.

## FEATURES

source	Location/Qualifiers
gene	1..31853 /organism="Caenorhabditis elegans" /db_xref="taxon:6239" /chromosome="IV" /clone="F44D12"
	join(3277..3413,3774..3951,3994..4239,4318..4603, 4647..4894,4945..5180,5225..5395,5449..5759,5805..6053, 6100..6802,6846..7020,7211..7306,7357..7425) /genes="F44D12.1"
	join(3277..3413,3774..3951,3994..4239,4318..4603, 4647..4894,4945..5180,5225..5395,5449..5759,5805..6053, 6100..6802,6846..7020,7211..7306,7357..7425) /gene="F44D12.1"
CDS	/note="Similarity to Mouse Tight junction protein ZO-1 (SW:Z01_MOUSE), contains similarity to Pfam domain: PF00595 (PDZ domain (Also known as DHR or GLGF)). Score=35.0, E-value=5.5e-07, N=2
	cDNA EST yk16d12.5 comes from this gene cDNA EST yk45f4.5 comes from this gene cDNA EST yk35f8.5 comes from this gene cDNA EST yk129c3.3 comes from this gene cDNA EST yk129c3.5 comes from this gene cDNA EST yk360b12.3 comes from this gene cDNA EST yk360b12.5 comes from this gene cDNA EST yk352d5.3 comes from this gene cDNA EST yk352d5.5 comes from this gene cDNA EST yk324a9.3 comes from this gene cDNA EST yk324a9.5 comes from this gene cDNA EST yk219a4.3 comes from this gene cDNA EST yk219a4.5 comes from this gene cDNA EST yk218f7.3 comes from this gene cDNA EST yk218f7.5 comes from this gene cDNA EST yk220g3.5 comes from this gene cDNA EST yk252d9.5 comes from this gene cDNA EST yk357f4.5 comes from this gene cDNA EST yk388f10.5 comes from this gene
	/codon_start=1 /protein_id="CAA92607.1" /db_xref="GI:3877158" /db_xref="SPTREMBL:Q20398"
gene	/translation="MEKEIREVEAKWMEKVRKASERDDIGKKIDFEFEKIDSLOIQV HRSETYKMEIEANKRLRQYEQALQRDYETMOERSIVLEENSRQNEERDLQKELEK MKEDSEVALKQLENTRLKLVNMEETAQANKEAARLEATEAQNGARITEERD EALRQCOQTQDAIVDWNTHSVQINLPFNKPLGIMLGGRDIDGSIHVGPYIVRQI AHGSPFDNLKLDHIMWINDISVTDMDERSVGMLSNCHHILVIRRSNCNKISDV CLPLNVLGSLNSGVFINSCPNGAASRLAPGVVHVMTPTVYDAKHAEMLIK SREPLVIGLQSTKRGDNGKDKHROTIFSRWFSRNGSDKERTVYVAKNIDRSNDQV LLRQSLRWPQASPMSPLRYGSLRAPYSSIDHTKLMLDALKRFRNAKTAHSE ISSSTAAPTALMSPSSVNEKOLVFAGTNEGVPVYVPKSTIFSTPTSPVRLIRQNS EQKRTMSQISVSHRGDGRFSNTTSDAPYSMHFTPTSTIMEGPHRRSAVYSP SHPVYPTTRDMSVMSVSSNSIRLPSTFSNQYNNSTCSLTGGSGVPPRHCISKD GSDFSITSTGSAHSAEGRSNKKYHLPNGNHSHESLVNRRSQPRLEVEPRSDVKLC GGNIGILAEKISDGLLEDLLISIDGNCVRNTTLECAMNTLSADNSELTLGLVODG GRLNRLRGADGDSFRLNIDRSMENDELKCKGCVFVDKTLMLGKTGRRAWK VDKRGQRHGAIPSTVTYQAIRANRYANFPFKAEVWEKLDTKVRPVLIRGAVY EPTQLMLVDESRESFVARESLASLDEVSALLKDKVLDSKQNDVDYILYHVSTA IMDTAQGLHCLVQVQSAIDRLKRCMFPILVKIRFQSVKQKDVNEHCIGEKISK EAKOLIEKDLKIEKLDGSGVTLVVPSHNNSFEMTHAVLQLKKIIEDEQKIVWQVK VDEE"
	complement(join(7556..7673,7717..7832,8096..8296, 8340..8510,8562..8718,8944..9111,9541..9638)) /gene="F44D12.2"
	/notes="cDNA EST yk169e12.3 comes from this gene cDNA EST yk367d4.3 comes from this gene cDNA EST yk367d4.5 comes from this gene cDNA EST yk553h9.3 comes from this gene cDNA EST yk553h9.5 comes from this gene" /codon_start=1 /protein_id="CAA92598.2" /db_xref="GI:5824507" /db_xref="SPTREMBL:Q20392"
gene	/translation="MOTILQLIVLFLFHEVNTLVTCRCEFGNECNFSRSTCRGD FCYHAQFYGNCMPPTIQKGVIGEVAPEGCRVHGHNVCMFCSDADYCNANTISLAD TATTILPVQCOPEKVNMPKPRWTKPCAANYCTFTEAKTQTEVNSANYTWSTKDCNK ENFDEFPTLTVFENFYGTCVWLNYGQDPTHCYGSDSLDTTLVDFDVTATWTECHVD YNPRPLPYKSGSSCICQCFISATSRGEVFRGCVNSVTEGATPLKIGTRYATGLE QWICDQSYCNADLKSAELSWPPELYLYRNISLRENFNVFYDARSSSSIFLAIPVF LINYGLFHF"
	join(12326..12380,12427..12481,12574..12895) /genes="F44D12.9"
	join(12326..12380,12427..12481,12574..12895) /gene="F44D12.9"
CDS	/codon_start=1 /protein_id="CAA92605.1" /db_xref="GI:3877156" /db_xref="SPTREMBL:Q20396"
	/translation="MTKHKMQEEGSPLSYFOYLHVEYNLCNGKRGOTTLLKNEELA ENSVGTGNSQIGLVTPRDWTGKRKESPTDRWLYVLIFTMHGCHLMSWNMTIAPO YIHDYWFNNTYQDSEMSIIGVTSQIPNVGIMILNTIVVMV"
	complement(15660..15989) /genes="F44D12.3" complement(15660..15989) /gene="F44D12.3"
gene	/note="contains similarity to Pfam domain: PF00635 (MSP Major sperm protein domain), Score=111.7, E-value=4.7e-30, N=1
	cDNA EST yk57e8.3 comes from this gene; cDNA EST yk102a1.3 comes from this gene cDNA EST yk57e8.5 comes from this gene; cDNA EST yk102a1.5 comes from this gene cDNA EST yk122a5.5 comes from this gene; cDNA EST yk149d3.5 comes from this gene cDNA EST yk151a2.5 comes from this gene; cDNA EST yk164a1.5 comes from this gene cDNA EST yk170h1.5 comes from this gene; cDNA EST yk179d6.5 comes from this gene cDNA EST yk189f1.5 comes from this gene; cDNA EST yk149d3.3 comes from this gene cDNA EST yk151a2.3 comes from this gene; cDNA EST yk164a1.3 comes from this gene cDNA EST yk170h1.3 comes from this gene; cDNA EST yk179d6.3 comes from this gene cDNA EST yk189f1.3 comes from this gene; cDNA EST yk190a10.3 comes from this gene cDNA EST yk92f10.3 comes from this gene; cDNA EST yk150a11.5 comes from this gene cDNA EST yk164f1.5 comes from this gene; cDNA EST yk378f3.3 comes from this gene cDNA EST yk378f3.5 comes from this gene; cDNA EST yk431f3.3 comes from this gene cDNA EST yk431f3.5 comes from this gene; cDNA EST yk365g9.3 comes from this gene cDNA EST yk365g9.5 comes from this gene; cDNA EST yk311h6.3 comes from this gene cDNA EST yk311h6.5 comes from this gene; cDNA EST yk299b2.3 comes from this gene cDNA EST yk299b2.5 comes from this gene; cDNA EST yk309f7.3 comes from this gene cDNA EST yk309f7.5 comes from this gene; cDNA EST yk303a11.3 comes from this gene cDNA EST yk303a11.5 comes from this gene; cDNA EST yk302c7.3 comes from this gene

gene











```
FEATURES
  source      Location/Qualifiers
  1..7218
    /organism="unknown"
BASE COUNT   1944 a 1491 c 1486 g      1929 t      368 others
ORIGIN

Query Match      4.9%; Score 68; DB 6; Length 7218;
Best Local Similarity 4.2%; Pred. No. 3.8e-07;
Matches 17; Conservative 234; Mismatches 149; Indels 0; Gaps 0;

Qy 907 ttgtctcaaaagtctctgtctcaattgtctgctgcatggtatccatgttcttc 966
|||| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1060 TTGCGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1119

Qy 967 aagattcactggcgcaccccgcttactcaaatgccataattcgtctgctctttc 1026
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1120 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1179

Qy 1027 attcattcttctctgcaactatcgtgtccagacgctgtcttctcttcttt 1086
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1180 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1239

Qy 1087 gagtctactgacatttgtgatggtggaattgccatgtcttcttccatcggtac 1146
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1240 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1299

Qy 1147 agcctctggcaatggatcacactccaaacgtcgtgccatctactactcaagattgcc 1206
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1300 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1359

Qy 1207 gctcagcttctcgttgcactcttatgttgcctctccacgctggtccgctt 1266
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1360 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1419

Qy 1267 gtattgagcactctgtggacaaagccaagtatcttataaa 1306
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1420 YYYYYYYYYYYYYYGGACCAAAATCTCTATCTCTTAA 1459

RESULT 12
CEK09A9
LOCUS      CEK09A9                      34017 bp      DNA      linear      INV 11-DEC-2001
DEFINITION Caenorhabditis elegans cosmid K09A9, complete sequence.
ACCESSION  Z79601
VERSION    Z79601.1 GI:3646501
KEYWORDS   WFG; Aspartyl beta hydroxylase like; NADH-ubiquinol oxidoreductase;
           Protein Kinase; RAS-related protein; Ubiquitin carboxyl-terminal
           hydrolase.
SOURCE     Caenorhabditis elegans.
           Caenorhabditis elegans.
           Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
           Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE  1 (sites)
AUTHORS   none.
TITLE     Genome sequence of the nematode C. elegans: a platform for
           investigating biology. The C. elegans Sequencing Consortium
           Science 282 (5396), 2012-2018 (1998)
JOURNAL   99069613
MEDLINE   99069613
REMARK    The C.elegans Sequencing Consortium.
REFERENCE  2 (bases 1 to 34017)
AUTHORS   Swinburne, J.
TITLE     Direct Submission
JOURNAL   Submitted (29-AUG-1996) Nematode Sequencing Project, Sanger
           Institute, Hinxton, Cambridge CB10 1SA, England and Department of
           Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
           jes@sanger.ac.uk or rw@nematode.wustl.edu
           On Sep 24, 1998 this sequence version replaced gi:1515146.
           Coding sequences below are predicted from computer analysis, using
           predictions from Genefinder (P. Green, U. Washington), and other
           available information.
           Current sequence finishing criteria for the C. elegans genome
           sequencing consortium are that all bases are either sequenced

unambiguously on both strands, or on a single strand with both a
dye primer and dye terminator reaction, from distinct subclones.
Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is not the entire insert of clone K09A9.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone K09A9 is at 35278 in
sequence Z79596.
The true right end of clone K09A9 is at 34017 in this sequence. The
true left end of clone K09E9 is at 24804 in this sequence. The true
right end of clone C02c6 is at 104 in this sequence. The start of
this sequence (1..104) overlaps with the end of sequence Z79596.
The end of this sequence (33914..34017) overlaps with the start of
sequence Z79602.
For a graphical representation of this sequence and its analysis
see: - http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?
name=K09A9
IMPORTANT: This sequence is NOT necessarily the entire insert of
the specified clone. It may be shorter because we only sequence
overlapping sections once, or longer because we arrange for a small
overlap between neighbouring submissions.
Location/Qualifiers
1..34017
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/chromosome="X"
/clone="K09A9"
complement(join(9270..9399,10952..11597,11646..11868,
12089..12263,12317..12450,12502..12642))
/gene="K09A9.5"
complement(join(9270..9399,10952..11597,11646..11868,
12089..12263,12317..12450,12502..12642))
/gene="K09A9.5"
/note="predicted using Genefinder
Similarity to Bovine NADH-ubiquinone oxidoreductase 49 KD
subunit (SW:NUCM_BOVIN), contains similarity to Pfam
domain: PF00346 (Respiratory-chain NADH dehydrogenase, 49
Kd subunit). Score=595.8, E-value=6.4e-215, N=1
CDNA EST EMBL:Z14772 comes from this gene
CDNA EST EMBL:M88875 comes from this gene
CDNA EST EMBL:M89294 comes from this gene
CDNA EST EMBL:T01723 comes from this gene
CDNA EST Yk81b6.3 comes from this gene
CDNA EST Yk81b6.5 comes from this gene
CDNA EST Yk68h5.3 comes from this gene
CDNA EST Yk68h5.5 comes from this gene
CDNA EST Yk70h1.5 comes from this gene
CDNA EST Yk70h1.3 comes from this gene
CDNA EST Yk501g3.3 comes from this gene
CDNA EST Yk485g3.3 comes from this gene
CDNA EST Yk485g3.5 comes from this gene
CDNA EST Yk454f8.3 comes from this gene
CDNA EST Yk454f8.5 comes from this gene
CDNA EST Yk424c6.3 comes from this gene
CDNA EST Yk424c6.5 comes from this gene
CDNA EST Yk418a5.3 comes from this gene
CDNA EST Yk418a5.5 comes from this gene
CDNA EST Yk229c10.3 comes from this gene
CDNA EST Yk229c10.5 comes from this gene
CDNA EST Yk409d1.5 comes from this gene
CDNA EST Yk530c9.3 comes from this gene
CDNA EST Yk532b10.3 comes from this gene
CDNA EST Yk596h9.3 comes from this gene
CDNA EST Yk652g6.3 comes from this gene
CDNA EST Yk680g8.3 comes from this gene
CDNA EST Yk530c9.5 comes from this gene
CDNA EST Yk562f5.5 comes from this gene
CDNA EST Yk675a3.5 comes from this gene"
/codon_start=1
/protein_id="CAB01886.1"
/db_xref="GI:3878424"
/db_xref="SWISS-PROT:Q93873"
```

/translation="MLGRKIAGTCLRANVPSVAATSTPATQTRNSHTIWYDPAKFER  
QFTGTTGLKMSERYSDFEKIKGLDLEKLAISDPMDSNDYSGKQREKNLENNILN  
FGPOHAAAGRLVLKLEGEVILKAPHIGLLRHAYSDPMDSNDYSGKQREKNLENNILN  
VSMNCQASLAVEKILGIDIPTRAKYIRTLGELPRIONHLMGITHALDVCAMP  
FFWFERELKFEPSERGARMHANVVRGGVAVMDLPILGMDIDWAKIKFFPRIDE  
LEMLTRNWKARTIDIGLVSADALNMGFSGVMVRGSGIKQDVRKTEPYDAYDME  
FDVPIGTGCDYRCLRCPEEATVPIEAPKGFVYLAVDGKPKYRCITRAPFAH  
NMESLIHHFFFTFGVQPPGATVPIEAPKGFVYLAVDGKPKYRCITRAPFAH  
LAATHVCYNSLLADIIVAVIGTMDIVFGEVDR"  
join(13620..13793,13874..14074,14437..14532,14631..14697,  
14741..14841,14889..15017,15106..15351,15397..15632,  
15939..16205,16346..16496)  
/gene="K09A9.4"  
join(13620..13793,13874..14074,14437..14532,14631..14697,  
14741..14841,14889..15017,15106..15351,15397..15632,  
15939..16205,16346..16496)  
/note="Similarity to Yeast ubiquitin carboxyl-terminal  
hydrolase (SW:UBPD\_YEAST), contains similarity to Pfam  
domain: PF00442 (Ubiquitin carboxyl-terminal hydrolases  
family 2). Score=55.1, E-value=4.9e-13, N=1; PF00443  
(Ubiquitin carboxyl-terminal hydrolase family 2),  
Score=12.8, E-value=0.00029, N=1  
cDNA EST EMBL:Z14811 comes from this gene  
cDNA EST yk527f6.5 comes from this gene  
/codon\_start=1  
/protein\_id="CAB01885.1"  
/db\_xref="GI:3878423"  
/db\_xref="SP:PREMBL:Q93875"  
/translation="MADSAGRDDGMSDSQPSQNETKEEGILDESSSESGETGPD  
TKDLDENNPNRMLPKMESCRRKHVKIDNPDLSQLQKLLKCMTKCHDSVSFSCAH  
QDCINLSNCCCKHLLPHMAVHNHPILYHDKRSILCTRCMVYPPYVRRFTHTHF  
STSDKQDETKEPTEAFRLGLVLFNGNTCYMNAVLQLLGHGSPYVPIELIELIPGG  
WSNCSHDIPKTAQMALDRNYSDEPLVYLSWKIITQODASEFMEMLDJLDLDDLK  
TCSEYHRTNLIENGEMDIYATMANPHYKTAISAFQVLENOICHCGRFSRTI  
ENFDLSLIVGENEFEMASKLDDPKAPCTSSNIEDGDDPGFVHQAFNKTSL  
SLCDRFQNTLDDNDSYSCKEVLVDATKSTKAHLRPEVILQLKRRFTWFGSK  
VGVKEPPLSQDFGRWYTSCEPALYDLGVFVHGSRLEFHYVSVCRHEQDNQCCR  
HVRGSEWRKGGFVYGGORNAKRLFAGRHVRGSTPHYSFGYMLRL"  
join(23324..23500,23815..23981,24388..24482,24569..25081,  
25131..25216,25261..25575)  
/gene="K09A9.3"  
join(23324..23500,23815..23981,24388..24482,24569..25081,  
25131..25216,25261..25575)  
/note="K09A9.3"  
/note="predicted using GeneFINDER  
contains similarity to Pfam domain: PF01733 (Nucleoside  
transporter). Score=501.6, E-value=1.9e-147, N=1  
cDNA EST EMBL:M79812 comes from this gene  
cDNA EST yk106d4.3 comes from this gene  
cDNA EST yk78h6.3 comes from this gene  
cDNA EST yk111h7.3 comes from this gene  
cDNA EST yk116g9.3 comes from this gene  
cDNA EST yk78h6.5 comes from this gene  
cDNA EST yk106d4.5 comes from this gene  
cDNA EST yk111h7.5 comes from this gene  
cDNA EST yk116g9.5 comes from this gene  
cDNA EST yk143f7.5 comes from this gene  
cDNA EST yk183c7.5 comes from this gene  
cDNA EST yk183c7.3 comes from this gene  
cDNA EST yk143f7.3 comes from this gene  
cDNA EST yk488b8.3 comes from this gene  
cDNA EST yk488b8.5 comes from this gene  
cDNA EST yk247h5.3 comes from this gene  
cDNA EST yk247h5.5 comes from this gene  
cDNA EST yk215h7.5 comes from this gene  
/codon\_start=1  
/protein\_id="CAB01882.1"  
/db\_xref="GI:3878420"  
/db\_xref="SP:PREMBL:Q93871"  
/translation="MSRAVELQFLNEAKKVEDAEVEETPEPEDKGNLVYIILLHGI  
GLTLPWNMLTISDYDFESYKMLANSTIDMTGKVDGTPVYSNFSQFTIASQVFN  
LLNLLNIFIVVKGGLAGRITVGLSIVAVCVITMTFIYVITSTWLTGFTLTITII

VLNANGVYQNSIFGLASELFFKYNTNAVIIGNNLCGTFVTLLSMSTKAVTRNILDRSF  
AYPSALLNVFVCFISHLIKKQRFYQFSTRAERQAKNEEADNGCKWANYIATEFK  
EAPOLLNVLFVFTLSIFPGVMMYVKDEKKGCTYDFPLPQNYFMDVTFLOQNVFA  
FIQSTVAGRKOMPAPKMLIPVYLRLYIPFFICNVLPETRLPLVPFESFWLFLIVA  
ASMSFGSGYSGLSMMYTSKTVDPDSKAQVAMGAGFLISGIVSGLIFTWVWIKFVVA  
\*  
complement(join(27534..27696,27760..27878,28371..28615,  
28762..28867))  
/gene="K09A9.2"  
complement(join(27534..27696,27760..27878,28371..28615,  
28762..28867))  
/note="Similarity to Rat RAS-related protein RAB-14  
(SW:RB14\_RAT), contains similarity to Pfam domain: PF00071  
(RAS family). Score=319.6, E-value=1.2e-92, N=1  
cDNA EST yk115b5.3 comes from this gene  
cDNA EST yk115b5.5 comes from this gene  
cDNA EST yk479h5.3 comes from this gene  
cDNA EST yk479h5.5 comes from this gene  
cDNA EST yk293f2.3 comes from this gene  
cDNA EST yk293f2.5 comes from this gene  
cDNA EST yk197e3.5 comes from this gene  
cDNA EST yk529d11.5 comes from this gene  
Query Match 4.7%; Score 65.8; DB 3; Length 34017;  
Best Local Similarity 51.0%; Pred. No. 1.7e-06;  
Matches 182; Conservative 0; Mismatches 172; Indels 3; Gaps 1;  
Qy 978 gccagacacccggttacctcaaatggccataattggcgctgctcttttcattccattctt 1037  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 25260 GCCGCCGCCGAAACAGCTCTGGATGCCACCTGCTCGCTCTCTCATCATCCCATCTT 25319  
Qy 1038 cttcttgcgaactcgtgtccagcgcgtgtattctctctttttcttgagtcactga 1097  
||||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 25320 CATCTTGCACACTACCTCCCTCGACCCGACACTCCCGGCTCTTCTTCGAATCCACGTG 25379  
Qy 1098 cattttgtgattggtggaattgccaattctttttcacatggataccctcagcgtctggc 1157  
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 25380 GCTTTTCATCATCTTGGCGCATCGATGAGCTTCGGAAGTGGATACTTCTCAGGACTCTC 25439  
Qy 1158 aatgggatacactccaacacgtcgtgccaatctcactactcaagattgcccgtcagctttc 1217  
||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 25440 CATGATGTACACTTCAAAAGACTGTCGATCCATCAAAAGCTCAGGTGGTGAATGATGCG 25499  
Qy 1218 cgttgcactcttattggtgaccttcacccggtggtggtggtggtggtggtggtggtggt 1277  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 25500 TGGATTTTTCCTCATTTCCAGGAATCGTCTCAGGCTCATTTTTCATATGGTCATCAAG-- 25557  
Qy 1278 cttcgtgacacagccaagtattctataaatatttatagcattagatgatacttggta 1334  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 25558 -TTTGTGCTACTGCCATAAATTTTCAAAACTTCTACATTAAAGACACCTCATTA 25613  
RESULT 13  
PFMAL3P7  
LOCUS 253305 bp DNA linear INV 22-MAY-2000  
DEFINITION Plasmodium falciparum MAL3P7, complete sequence.  
ACCESSION AL008974 AL008975 AL008981 AL008983 AL009015 AL010138  
AL010143 AL010146 AL010154 AL010157 AL010160 AL010165 AL010169  
AL010187 AL010189 AL010207 AL010208 AL010214 AL021885 AL021888  
AL139179 298556 298557 298558  
AL034559.4 GI:8052273  
KEYWORDS HMG; 40S Ribosomal protein S3A; acyl transferase;  
acylaminoacyl-peptidase; ATP-dependent RNA Helicase; cyclophilin;  
elongation factor; F49c12.11-like protein; Hesp-like domain  
protein; histone H2A variant; kinesin-related protein;  
N-acetylglucosamine-1-phosphate transferase; protein kinase;  
repeat; R-FA3 repeat; repl1; rep20; rifin; RNA-binding protein;  
stevor; T-complex protein 1 epsilon subunit; telomere; var.  
SOURCE malaria parasite P. falciparum.  
ORGANISM Plasmodium falciparum  
REFERENCE 1 (bases 1 to 253305) Haemosporida; Plasmodium.  
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

AUTHORS	Bowman, S., Lawson, D., Basham, D., Brown, D., Brown, D., Chillingworth, T., Churcher, C. M., Craig, A., Davies, R. M., Devlin, K., Feltwell, T., Gentles, S., Gilliam, R., Hamlin, N., Harris, D., Holroyd, S., Hornsby, T., Horrocks, P., Jagels, K., Jassal, B., Kyes, S., McLean, J., Moule, S., Mungall, K., Murphy, L., Oliver, K., Quail, M. A., Rajandream, M. A., Rutter, S., Skelton, J., Squares, R., Squares, S., Sulston, J. E., Whitehead, S., Woodward, J. R., Newbold, C. and Barrell, B. G.		
TITLE	The complete nucleotide sequence of chromosome 3 of Plasmodium falciparum		
JOURNAL	Nature 400 (6744), 532-538 (1999)		
MEDLINE	99376085		
REFERENCE	2 (bases 1 to 253305)		
AUTHORS	Oliver, K., Bowman, S., Harris, D., Lawson, D., Quail, M. and Barrell, B.		
JOURNAL	Unpublished		
REFERENCE	3 (bases 1 to 253305)		
AUTHORS	Lawson, D., Bowman, S. and Barrell, B.		
TITLE	Direct Submission		
JOURNAL	Submitted (17-DEC-1998) P.falciparum Genome Sequencing Consortium, The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK		
COMMENT	On May 14, 2001 this sequence version replaced gi:2982540 gi:2982541 gi:2982544 gi:2984453 gi:2984464 gi:2982551 gi:2982556 gi:2982558 gi:2982563 gi:2982564 gi:2982566 gi:2984491 gi:2982569 gi:2695974 gi:2894502 gi:2673766 gi:2894496 gi:2982577 gi:2894588 gi:2894589 gi:2982538 gi:2982539 gi:2894380 gi:4725992. For more information about this sequence or the Malaria Project, see <a href="http://www.sanger.ac.uk/Projects/P_falciparum">http://www.sanger.ac.uk/Projects/P_falciparum</a> .		
FEATURES	Location/Qualifiers		
Source	1..253305		
	/organism="Plasmodium falciparum"		
	/strain="3D7"		
	/db_xref="taxon:5833"		
	/chromosome="3"		
	/clone="MAL3P7"		
gene	458..4060		
	/gene="PFC0860w, MAL3P7.1"		
CDS	458..4060		
	/gene="PFC0860w, MAL3P7.1"		
	/note="PFC0860w (MAL3P7.1), Kinesin-related protein, len: 1200 aa; similarity: to kinesin-related protein.		
	D.melanogaster kinesin-like protein 67A (Tr:P91945) BLAST Score: 664, sum P(2) = 1.9e-67; 28% identity in 707 aa overlap, predicted using hexExon, Pfam: match to PF00225 kinesin, Kinesin motor domain Score 311.54"		
	/codon_start=1		
	/product="kinesin-related protein"		
	/protein_id="CAB39023.1"		
	/db_xref="GI:4493964"		
	/db_xref="SPTREMBL:O97277"		
	/translation="MRKTKNDLPFLNNYYNKEENNPIIICDERKERTKIYSTNH HTNNILYNQHTLTKICNITENNKAISNLOKNNILMPYSYMKQKGTIRETIKN YYPNINEGSTVENITNGEHTINGOYDALKNLSNNDYHQHNNIMNINKNKLFVL NNNNNNQDVNNENYPTISKNYDSIYIQTINKIINSEINNINNNDNNHH NYTNHNNNNNSPSSVTSKNNNNINTLNAYSNKVAIVRPIGESENIIV IFKNYLIIEKNEKCYLLSQKKQSTYVDFVDVNAEQEEVFTAKPLIPHFVK GINTCFAYGATSGKTYTMDLQKNGQIVQLSLELFTIINEKCRNIKVLSFLEV YNETIRDLGCKEKNTLEVOEVAEVKVSICETEIVNVEQALLINGVKNRKMSP T RANKVSRSRAILQIVYVNEILLDDMMNTISYKAKLCFVDLAGSERASATSKGERPKE GSYINQSLALALACINSIASNRNLSKVRIRKSLTHLLKNSLEGCLVMTIANINP SRTYSFQSNNTATFARNIKLCATQVTDNKESDIEILKKNENLOKQEDYTLGKY TNLKFFPIININQYKQISYKLIENISDNMSMELQDITMYDQLVKMKSDEYR KKVSDLDLYQEEKQFLNNLFDYFLEKNLVNTNSKDVNDNNKSLLEMIFFKHNEK VVENFLVNEKVDKNVNLGNVNVVDENVVNDKVLNNGVNVVDENVVDENVVM VDNVNVVDENVVDENVVDENVVDENVVDENVVDENVVDENVVDENVV AEGKSESHNNKNDKIDEDKDTIKDIHNNNSDNDDEYQSANSPEVDIVKKEKK KKTIPINNETKKTMTNGTDPIHKTPYDINIIGLNKEDYSKNSDNTNKNTEKKN EKKGEYFPFNHTDMGNSILYINNNVNSPSRPMKNAVKMLKGLNLTANFTL DDDT IKMNSNGKISDKHNMKNLNNNGKINDKSKCKNINNNNNNNNNNNNN NNNNNSSSSSGSKVDGINLANSNTNERLHTFGSVYSLNLANDEIKIETNKDEKN DIHLTSDTISKIHARDLLKENRKKLENFOENIKHEKDEVSLYVKKKKKKKK"		
	4028..4033		
	/gene="PFC0860w, MAL3P7.1"		
misc_feature	/note="potential splice donor at 3' end of PFC0860w may indicate splicing"		
	4057..4062		
	/note="potential splice donor at 3' end of PFC0860w may indicate splicing"		
gene	join(5457..5577,5689..5787,5923..6077,6189..6743)		
	/gene="PFC0865w, MAL3P7.2"		
CDS	join(5457..5577,5689..5787,5923..6077,6189..6743)		
	/gene="PFC0865w, MAL3P7.2"		
	/note="PFC0865w (MAL3P7.2), RNA-binding protein, len: 309 aa; Similarity eg to C.elegans RNA-binding protein (TR:Q18318) BLAST Score: 378, sum P(1) = 3.8e-35; 37% identity in 297 aa overlap, predicted using hexExon, Pfam: match to PF00076 rrm, RNA recognition motif. (a.k.a. RRM, RBD, or RNP domain) Score 58.01"		
	/codon_start=1		
	/product="putative RNA-binding protein"		
	/protein_id="CAB39067.1"		
	/db_xref="GI:4494008"		
	/db_xref="SPTREMBL:O97318"		
	/translation="MGLFDKIRNIEKLNKNEAEKLNIGNNDSDHWDQYRDSSYIYIGILD NLTREGDIVIVFSOYGPIDVNLVRDNETGKGCYCFLSYADQSTILAVDNFNGYKL LERPLVVDHILNLYRLPKKYLKADKNKKYKPGASGQIGIVYVNSEKLKSKVDKIK NKSNEERKKLLDEDEWALNFERSIKRDIISPIGHDEKSHRNGMEKEDEDEDD DDEDDDDSDVDIKYKRHKERKSLATTKYDKKHKRKSDDHDKHRRRHHSHREK EKDKKSHRRHKHSYDKYSRSRSYSTSSSTDR"		
gene	join(7801..7928,8084..8234,8392..8583)		
CDS	join(7801..7928,8084..8234,8392..8583)		
	/gene="PFC0870w, MAL3P7.3"		
	/gene="PFC0870w, MAL3P7.3"		
	/note="PFC0870w (MAL3P7.3), putative elongation factor, len: 181 aa; Similarity to P.falciparum elongation factor 1-beta, (AAF27524) (276 aa), fasta scores: opt: 366, E(1): 1.1e-15, 53.2% identity in 141 aa overlap, revised: shortened exon 2, Pfam: match to PF00736 EF1BD, EF-1 guanine nucleotide exchange domain Score 61.10"		
	/codon_start=1		
	/product="putative elongation factor"		
	/protein_id="CAB39068.2"		
	/db_xref="GI:8052274"		
	/db_xref="SPTREMBL:O97319"		
	/translation="MANTYDELYPLSYIQLQNEGGNTSKIDQANTKKPKKKEVINKS SLIIDIKPTEGTDLDVLRKNTMDEGTWGAHKKTPTFAFLKQLQVSCVIVDDL VNTDELIIYENTGLDNEQLQKKQMDDEYDEDDDEIGSLVQSAEIIISPNKL" 8235..8240		
misc_feature	/gene="PFC0870w, MAL3P7.3"		
	/note="revised splice donor sequence for exon 2 of PFC0870w"		
misc_feature	8580..8585		
	/note="potential splice donor sequence may indicate further splicing of PFC0870w"		
gene	join(11431..20416,20578..20690)		
CDS	/gene="PFC0875w, MAL3P7.4"		
	join(11431..20416,20578..20690)		
	/gene="PFC0875w, MAL3P7.4"		
	/note="PFC0875w (MAL3P7.4), hypothetical protein, len: 3032 aa, revised: added 3' exon, possibly spliced at 5' end"		
	/codon_start=1		
	/product="hypothetical protein, PFC0875w"		
	/protein_id="CAB39024.2"		
	/db_xref="GI:8052275"		
	/db_xref="SPTREMBL:O97278"		
	/translation="MKDFLSYAENNDLNIKLYNSEEDCLNKLKAYLELSEDFKNVN SLENDLILHENDKNSLYNLHTNINILKSSFNINDLKLYVNFNFENFNKIKI NPHLLKVVYKTEFYKFIKDIKLNHNHYQYFNDDRDKLFYVFNKNVLETYKSC GLLSVNVNYKKQKNVYKFLGSPQSEGFKGLYKNAPLYKNKYGYGNPFHL YYPVYVLESOKKIKNEKNEKNEKNEKNEKNEKNEKNDKILSSFTNNVWVK NGIOSYINSNSNDQKKNNNNNYYYNNPDGLITVYKIRYGVQVALLNSNEK YFFNDIINLKNQNFNFVDDLSFNLYFNWYVSEFVLEFYQSFILNADILK KNNLLTPRYDEINELNEREITMEYKKTDEKEYOVNNEYINENVGYDEKGYKTEK LNDFFIYKMPKSMKINAFDTFEKNIIRVVVFLCVCLFIVNICDINKERKINLENFL CLKIKINYYYYYSLWLLFYFILFYIIFLVVYVYVYKLMNYFILFLFYFLMNS		





**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 23, 2002, 19:05:27 ; Search time 431.47 Seconds  
(without alignments)  
5523.159 Million cell updates/sec

Title: US-08-816-011f-36  
Perfect score: 1388  
Sequence: 1 atggaataatcaacagatc.....ttattataaaaaaaaaaaaaa 1388

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 25 summaries

Database : N\_Geneseq\_032802.\*  
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1388	100.0	1394	22 AAS12162	Caenorhabditis ele
2	1386.4	99.9	1388	17 AAT18168	CORK potassium cha
3	90.8	6.5	2060	23 ABL15935	Drosophila melanog
4	90.8	6.5	4060	23 ABL15934	Drosophila melanog
5	43.8	3.2	2354	23 ABL22074	Drosophila melanog
6	43	3.1	8237	22 AAS46801	Tumour suppressor
7	41	3.0	7657	22 AAS45477	Chemically pretrea
8	41	3.0	7657	24 ABL34022	Human immune syste
9	40.8	2.9	15732	22 AAS45389	Chemically pretrea

c	10	39.4	2.8	362	22	AAF65894	Novel human polynu
c	11	39.4	2.8	18218	24	ABL33948	Human immune syste
c	12	39.2	2.8	440	18	AAT67765	H. pylori secreted
c	13	39.2	2.8	440	18	AAT77445	H. pylori secreted
c	14	39.2	2.8	14112	22	AAS46477	Tumour suppressor
c	15	39.2	2.8	14112	24	ABL33442	Human immune syste
c	16	38.6	2.8	310	22	AAH71505	Human cervical can
c	17	38.6	2.8	1558	17	AAT28255	Survival motor neu
c	18	38.6	2.8	1560	17	AAT18828	Human survival mot
c	19	38.6	2.8	1582	17	AAT28259	Human survival neu
c	20	38.6	2.8	1582	17	AAT18831	Human survival mot
c	21	38.4	2.8	2439	20	AAH80667	Clone amb56_3 enco
c	22	38.4	2.8	2439	22	AAS59217	Human cDNA encodin
c	23	38.4	2.8	2439	24	ABA90886	Human polynucleoti
c	24	38.4	2.8	5945	24	ABL32084	Human immune syste
c	25	38.4	2.8	6125	24	ABL33613	Human immune syste
c	26	38.4	2.8	10732	21	AAA10594	Gene encoding a su
c	27	38	2.7	3499	21	AAZ55832	Staphylococcus aur
c	28	38	2.7	4709	18	AAV74898	Staphylococcus aur
c	29	38	2.7	6123	24	ABL33036	Human immune syste
c	30	37.8	2.7	8305	24	ABL33568	Human immune syste
c	31	37.8	2.7	12507	24	ABL32298	Human immune syste
c	32	37.8	2.7	15674	24	ABL32363	Human immune syste
c	33	37.8	2.7	15674	24	ABL34477	Human metastasis a
c	34	37.6	2.7	353	22	AAH14789	Human breast cance
c	35	37.4	2.7	422	22	AAH15924	Human breast cance
c	36	37.4	2.7	5504	24	AAS61308	Human gene regulat
c	37	37.4	2.7	7133	22	AAS46388	Tumour suppressor
c	38	37.4	2.7	7990	24	ABL32159	Human immune syste
c	39	37.4	2.7	12507	24	ABL32299	Human immune syste
c	40	37.2	2.7	6056	24	ABL33026	Human immune syste
c	41	37	2.7	2079	21	AAA26360	Human secreted pro
c	42	37	2.7	2285	22	AAH26734	Human breast cance
c	43	37	2.7	6794	24	AAS61175	Human gene regulat
c	44	37	2.7	8201	24	ABL32307	Human immune syste
c	45	36.8	2.7	19380	24	AAS61427	Human gene regulat

ALIGNMENTS

RESULT 1  
AAS12162  
ID AAS12162 standard; cDNA; 1394 BP.

XX AAS12162;

XX 21-NOV-2001 (first entry)

XX Caenorhabditis elegans potassium ion channel CORK DNA.

XX Transmembrane potassium ion channel protein; inward potassium flux; ss;  
XX pest control; membrane potential; pesticide; antihelminthic; nematode;  
XX insect; CORK.

XX Caenorhabditis elegans.

XX Key Location/Qualifiers

XX CDS 1..1305

XX polyA\_signal /product= "C. elegans CORK protein"

XX 1359..1364 /tag= a

XX WO200161006-A2. /tag= b

XX PD 23-AUG-2001.

XX PF 14-FEB-2001; 2001WO-US04680.

XX PR 15-FEB-2000; 2000US-0503849.

XX PA (BADI ) BASF CORP.



FT CDS 1..1305  
FT /\*tag= a  
FT /product= CORK potassium channel  
FT 1359..1364  
FT /\*tag= b

XX  
FN W09613520-A1.  
PD 09-MAY-1996.  
XX

PF 25-OCT-1995; 95WO-US14364.  
XX  
PR 31-OCT-1994; 94US-0332312.  
XX  
PA (AMCY ) AMERICAN CYANAMID CO.  
XX  
PI Pausch MH, Price LA;  
XX  
XX  
DR WPI; 1996-239450/24.  
DR P-PSDB; AAR92315.  
XX

PT Potassium channel genes from Drosophila melanogaster and  
PT Caenorhabditis elegans - useful in assaying substances to determine  
PT effects on cell growth, and in inhibiting nematode and insect pests  
XX  
PS Claim 13; Fig 9; 79pp; English.

XX The sequence encodes potassium channel CORK protein from  
CC Caenorhabditis elegans, with 2 pore-forming domains between  
CC transmembrane helix domains. The sequence has been isolated by  
CC complementation of the potassium-dependent phenotype of  
CC Saccharomyces cerevisiae CY162 (trk1-delta) on low-potassium medium.  
CC A consensus polyadenylation site is found in the 3'-untranslated  
CC sequence, and is followed by a tract of 15 consecutive A residues.  
CC The DNA may be inserted in a vector and expressed in a host cell to  
CC assay substances to determine effects on cell growth. Potassium-  
CC agonists or potassium-antagonists identified by this method may be  
CC used as nematocides, anthelmintics or in therapy of cardiac  
XX disorders, etc.

SQ Sequence 1388 BP; 341 A; 318 C; 269 G; 460 T; 0 other;

Query Match 99.9% Score 1386.4; DB 17; Length 1388;  
Best Local Similarity 99.9% Pred. No. 0;  
Matches 1387; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 atggttaataacacgctgaacacccatgccgttgagcagggaagcattcccaagagac 60  
Db 1 atggttaataacacgctgaacacccatgccgttgagcagggaagcattcccaagagac 60  
Qy 61 aagtacaattgtctactggtcgtcattctgtgtgattcgaggtctctgccaagg 120  
Db 61 aagtacaattgtctactggtcgtcattctgtgtgattcgaggtctctgccaagg 120  
Qy 121 aatagtctactatgcgcctcgtgattattgtgaattattggttcaaacggatggc 180  
Db 121 aatagtctactatgcgcctcgtgattattgtgaattattggttcaaacggatggc 180  
Qy 181 gtggagacatgtattcgaagaattcattggtattgagattggttcacaacttcca 240  
Db 181 gtggagacatgtattcgaagaattcattggtattgagattggttcacaacttcca 240  
Qy 241 aacgcaagcattaatgtttcaacccgtctcattattgtgtggtccctgatctaccgc 300  
Db 241 aacgcaagcattaatgtttcaacccgtctcattattgtgtggtccctgatctaccgc 300  
Qy 301 gtctttgctcgtttgttcaacatcgtcaacccgtgaacattcattctcctcgtcatt 360  
Db 301 gtctttgctcgtttgttcaacatcgtcaacccgtgaacattcattctcctcgtcatt 360  
Qy 361 gttctggagccactgaagattccatgctcctggttttctggttaactcttggaaatggcg 420  
Db 361 gttctggagccactgaagattccatgctcctggttttctggttaactcttggaaatggcg 420

Db 361 gttctggagccactgaagattccatgctcctggttttctggttaactcttggaaatggcg 420  
Qy 421 acttcaatcaatttttagcaatgggctatatgaaactcggtttttaggttggatggcgat 480  
Db 421 acttcaatcaatttttagcaatgggctatatgaaactcggtttttaggttggatggcgat 480  
Qy 481 ttcccgacacactacattggcgctctcttattgattggaacacacatttgcggttgcgtgata 540  
Db 481 ttcccgacacactacattggcgctctcttattgattggaacacacatttgcggttgcgtgata 540  
Qy 541 acggttggaataatcggagtgacctattttctgaatgatgagcctaaacttttgcacatc 600  
Db 541 acggttggaataatcggagtgacctattttctgaatgatgagcctaaacttttgcacatc 600  
Qy 601 gctatttcggcactatcgttggtgatcctctgtgtgctgaattgcaacttttcttattc 660  
Db 601 gctatttcggcactatcgttggtgatcctctgtgtgctgaattgcaacttttcttattc 660  
Qy 661 aaaaagcaagattttcacactatcacccatcaaaaaggaattggatttcgcgaaaaggcg 720  
Db 661 aaaaagcaagattttcacactatcacccatcaaaaaggaatttcgcgaaaaggcg 720  
Qy 721 gaaacggacagacgctccatccatttttggaccacatttcacaaactgttatggcaca 780  
Db 721 gaaacggacagacgctccatccatttttggaccacatttcacaaactgttatggcaca 780  
Qy 781 ctcttcaatgtttgtctgtcttgcgttactctcaaatcttccctgttatgatgacc 840  
Db 781 ctcttcaatgtttgtctgtcttgcgttactctcaaatcttccctgttatgatgacc 840  
Qy 841 gtaccactcgtggagattccggcttctcaacaaaattatgtctgaaacgatgaaatc 900  
Db 841 gtaccactcgtggagattccggcttctcaacaaaattatgtctgaaacgatgaaatc 900  
Qy 901 tacatttgcacaaagtctcgtcttctcaatttgcgtgctgagtgatggatccatgatt 960  
Db 901 tacatttgcacaaagtctcgtcttctcaatttgcgtgctgagtgatggatccatgatt 960  
Qy 961 gcttccaagattcactggtccgacacccggttacctcaaatgttgcataatcttgcgtgct 1020  
Db 961 gcttccaagattcactggtccgacacccggttacctcaaatgttgcataatcttgcgtgct 1020  
Qy 1021 ctttcttccattcttcttcttctgcaactatcgtgtccagacgctgcttcttcttct 1080  
Db 1021 ctttcttccattcttcttcttctgcaactatcgtgtccagacgctgcttcttcttct 1080  
Qy 1081 ttctttgagtcactgacatttttggattggtggaattgccaatcttcttcttccatgga 1140  
Db 1081 ttctttgagtcactgacatttttggattggtggaattgccaatcttcttcttccatgga 1140  
Qy 1141 tacctcagcgtctggtgcaatgggatacacctccaaacgctgctccatctcaactcaaga 1200  
Db 1141 tacctcagcgtctggtgcaatgggatacacctccaaacgctgctccatctcaactcaaga 1200  
Qy 1201 ttgctgctcagcttccgcttcttctgcaacttattggttgccttctcaccggtgctgctg 1260  
Db 1201 ttgctgctcagcttccgcttcttctgcaacttattggttgccttctcaccggtgctgctg 1260  
Qy 1261 cccgtgttattgacacactcgtggacaagccaattcttataataattttatagcattta 1320  
Db 1261 cccgtgttattgacacactcgtggacaagccaattcttataataattttatagcattta 1320  
Qy 1321 gagtatactgttatgttcttcttataagcttggaataaaaaataattataaaaaa 1380  
Db 1321 gagtatactgttatgttcttcttataagcttggaataaaaaataattataaaaaa 1380  
Qy 1381 aaaaaaa 1388  
Db 1381 aaaaaaa 1388

RESULT 3  
ABLI5935



XX	Drosophila melanogaster expressed polynucleotide SEQ ID NO 42284.
DE	
XX	Drosophila; developmental biology; cell signalling; insecticide;
KW	pharmaceutical; gene; ss.
KW	
OS	Drosophila melanogaster.
XX	
PN	WO200171042-A2.
XX	
PD	27-SEP-2001.
XX	
XX	23-MAR-2001; 2001WO-US09231.
PF	
XX	
XX	23-MAR-2000; 2000US-191637P.
PR	11-JUL-2000; 2000US-0614150.
XX	
PA	{PEKE } PE CORP NY.
XX	
PI	Venter JC, Adams M, Li PWD, Myers EW;
PI	
DR	WPI; 2001-656860/75.
DR	P-PSDB; ABB71831.
XX	
XX	New isolated nucleic acid detection reagent for detecting 1000 or more
PT	genes from Drosophila and for elucidating cell signalling and cell-cell
PT	interactions -
XX	
PS	Claim 1; SEQ ID NO 42284; 2lpp + Sequence Listing; English.
XX	
CC	The invention relates to an isolated nucleic acid detection reagent
CC	capable of detecting 1000 or more genes from Drosophila. The invention
CC	useful in developmental biology and in elucidating cell signalling and
CC	cell-cell interactions in higher eukaryotes for the development of
CC	insecticides, therapeutics and pharmaceutical drugs. The invention
CC	discloses genomic DNA sequences (AB116176-AB130511), expressed DNA
CC	sequences (AB101840-AB116175) and the encoded proteins
CC	(ABBS7737-ABBT2072).
CC	The sequence data for this patent did not form part of the printed
CC	specification, but was obtained in electronic format directly from WIPO
CC	at <a href="http://ftp.wipo.int/pub/published_pct_sequences">ftp.wipo.int/pub/published_pct_sequences</a> .
XX	
SQ	Sequence 4060 BP; 1109 A; 951 C; 864 G; 1136 T; 0 other;

```

Query Match      6.5%; Score 90.8; DB 23; Length 4060;
Best Local Similarity 45.4%; Pred. No. 8.2e-14;
Matches 556; Conservative 0; Mismatches 632; Indels 36; Gaps 5;

QY 47 catttccaagagacaagtacaataattctactaggctgcgtcattcttgttggattcggag 106
   || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1506 cagcgccaaagataattcctatcgtgtctttattctcgtcgacgcgtgggca 1565
   || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

QY 107 ttcttgcgcatggaatgttctactactcgcgcc-----tgagttattatgtgaatt 160
   || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1566 cgctgatgcgatgaaacatgttcatcagcgccaagtctatttcgaggaactcaagttcg 1625
   || |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

QY 161 attggttcaaacggatggcgtggagacatggtattcgaagaattcatggattcttga 220
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1626 gaccaaacacacgctggccaggaagtgtaccgcaccatttcatgcataaatatgg 1685
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

QY 221 cgattggctcaaaccttcaaacgcaagcattaattgtttcaacctgttctctattattg 280
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1686 gcttgcctgcagattctcaacctgttctcaactggcttcaactggcgaacatttgcacattg 1745
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

QY 281 ctggctccctgatctacccggtcttttgcctcgggttttgcctcaaacatcgtcaacctgacaa 340
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1746 gcggcgacactgaccaccgaatgctcacagcatcattctcgaaatggtcattctgtgg 1805
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

QY 341 tcattctcatctcgtcatgttcttggagccactgaagattccatctcctgggttttct 400
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 1806 ttaccattatttggccatcctcg-----actcctcgcagctggcgggcgctctctct 1859
   |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

```

RESULT	5	
ABL22074/c		
ID	ABL22074	standard; DNA; 2354 BP.
XX		
XX	ABL22074;	
XX		
XX		
DT	26-MAR-2002	(first entry)
XX		
DE	Drosophila melanogaster	genomic polynucleotide SEQ ID NO 17695.
XX		
KW	Drosophila;	developmental biology; cell signalling; insecticide;
KW	pharmaceutical;	gene; ds.

```

XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-658860/75.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Claim 1; SEQ ID NO 17695; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 2354 BP; 761 A; 460 C; 417 G; 716 T; 0 other;

Query Match 3.2%; Score 43.8; DB 23; Length 2354;
Best Local Similarity 76.1%; Pred. No. 0.2;
Matches 54; Conservative 0; Mismatches 17; Indels 0; Gaps 0;

Qy 1318 tttagagtactgtgttatgtgtttttatttaagctgtggaataaataattataaa 1377
Db 412 TAAGATTATATTATTAAAGATATTATTATTATTAACAGAGAAAAAAGATTATTAAAA 353

Qy 1378 aaaaaaaaaa 1388
Db 352 AGAAAAAAAAAA 342

RESULT 6
AAS46801
ID AAS46801 standard; DNA; 8237 BP.
XX
XX AAS46801;
XX
XX 18-DEC-2001 (first entry)
XX
XX Tumour suppressor gene derived chemically modified sequence #527.
XX
XX Human; tumour suppressor gene; oncogene; antitumour; cytostatic;
XX cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;
XX cytosine methylation; ds.
XX
XX Homo sapiens.
XX
XX WO200168912-A2.
XX
XX 20-SEP-2001.
XX
XX 15-MAR-2001; 2001WO-EP02955.

```

```

XX 15-MAR-2000; 2000DE-1013847.
XX 06-APR-2000; 2000DE-1019058.
XX 07-APR-2000; 2000DE-1019173.
XX 30-JUN-2000; 2000DE-1032529.
XX 01-SEP-2000; 2000DE-1043826.
XX PA (EPIC-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX WPI; 2001-602752/68.
XX
XX Fragments of chemically modified genes associated with tumour suppressor
XX genes and oncogenes, useful in designing primers and probes for
XX analysing diseases associated with cytosine methylation state e.g.
XX cancer -
XX
XX Claim 1; SEQ ID No 527; 27pp; English.
XX
XX The invention relates to a nucleic acid comprising a sequence of 18
XX bases, of a segment of chemically pretreated DNA (CP DNA) e.g. with
XX bisulphite, of genes associated with tumour suppression and
XX oncogenes having a sequence taken from 536 (actually 533 since
XX numbers 408, 458 and 500 are missing from the sequence listing) sequences
XX (SS) and sequences complementary to (SS). The nucleic acid may be a
XX peptide nucleic acid-oligomer (PNA) of at least 9 nucleotides and may
XX form part of a set of probes for detecting the cytosine methylation state
XX and/or single nucleotide polymorphisms and also to be used in an
XX array for analysing diseases associated with CpG dinucleotides e.g.
XX cancers and tumours. The probes can also be used in a method for
XX ascertaining genetic and/or epigenetic parameters for the diagnosis
XX and/or therapy of existing diseases or the predisposition to specific
XX diseases, by analysing cytosine methylations. The parameters may be
XX compared to another set of genetic and/or epigenetic parameters, the
XX differences serving as basis for diagnosis and/or prognosis events which
XX are disadvantageous to patients. The present sequence is one of the
XX 533 genomic sequences derived from tumour suppressor genes and
XX oncogenes.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 8237 BP; 2580 A; 53 C; 1572 G; 4032 T; 0 other;

Query Match 3.1%; Score 43; DB 22; Length 8237;
Best Local Similarity 58.0%; Pred. No. 0.53;
Matches 76; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Qy 1256 tctggcccgctgttatttgagcacttcgtggagcaagccaagtcattctataattattatag 1315
Db 5363 tctagatagttattataaaattatgttggagttatagattataaaataataaa 5422

Qy 1316 cattagagtatactgtttatgtttttttatttaagctgtggaataaataattataa 1375
Db 5423 aagaataatatttagaataatgttaattttgtttattatgttgaataattgttttaa 5482

Qy 1376 aaaaaaaaaa 1386
Db 5483 aataaaataaa 5493

RESULT 7
AAS45477
ID AAS45477 standard; DNA; 7657 BP.
XX
XX AAS45477;
XX
XX 18-DEC-2001 (first entry)
XX
XX Chemically pretreated complementary DNA associated with cell cycle #91.

```









Qy 962 cttccaagattcactgcccgcacaccccgtaactcaaatgttgccataatcttgctgctc 1021  
 :: | ||| | |||::: | || |||| ::| | ||| | |||| ||  
 Db 418 YVACAAAGCGCGCGSYVYVYAAATTTCCCTCTCKTKTCTAAATTTGGCGTCTT 359  
 Qy 1022 ttttcattccattctctctctgcgaactatcgctgcagacgctgcttatctctgtt 1081  
 ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |  
 Db 358 TTTCAAGCTCTTTTTTTCCTTTTCCAAATAGCGTTTTTTCATCTCTTACAAATTTCTTGT 299  
 Qy 1082 tcttgagtcta 1093  
 || | ||||  
 Db 298 GCTGCGCTTCTA 287

RESULT 13  
 AAT77445/C  
 ID AAT77445 standard; DNA; 440 BP.  
 XX  
 AC AAT77445;  
 XX  
 DT 11-AUG-1997 (first entry)  
 XX  
 DE H. pylori secreted or periplasmic protein ORF 24256572.aa.  
 XX  
 KW Transmembrane; cytoplasmic; cell envelope; flagella; transport;  
 KW secreted; periplasmic; chronic gastritis; duodenal ulcer disease;  
 KW activator; inhibitor; bacterial life cycle; vaccine; immunise;  
 KW detection; antisense; inhibition; ds.  
 XX  
 OS Helicobacter pylori.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 1..440  
 FT /\*tag= a  
 FT /transl\_except= (pos: 43..45, aa: Xaa)  
 FT /transl\_except= (pos: 379..381, aa: Xaa)  
 FT /transl\_except= (pos: 197..199, aa: Xaa)  
 FT /transl\_except= (pos: 200..202, aa: Xaa)  
 FT /transl\_except= (pos: 418..420, aa: Xaa)  
 FT /transl\_except= (pos: 421..423, aa: Xaa)  
 FT /transl\_except= (pos: 424..426, aa: Xaa)  
 FT /transl\_except= (pos: 427..429, aa: Xaa)  
 FT /note= "Xaa = unknown"  
 XX  
 XX WO9719098-A1.  
 XX  
 XX 29-MAY-1997.  
 XX  
 XX 15-NOV-1996; 96WO-US18542.  
 XX  
 XX 17-NOV-1995; 95US-0561469.  
 XX  
 XX (ASTR ) ASTRA AB.  
 XX  
 XX Smith DH;  
 XX  
 XX WPI; 1997-298052/27.  
 XX P-PSDB; RAN24627.  
 XX  
 XX Helicobacter pylori nucleic acid sequences and related proteins -  
 XX used for diagnostics and therapeutics  
 XX  
 XX Claim 1; Page 96-97; 1481pp; English.  
 XX  
 XX This sequence encodes a H. pylori protein likely to be secreted or  
 XX periplasmic.  
 XX Helicobacter pylori has been strongly linked to chronic gastritis and  
 XX duodenal ulcer disease. The nucleic acid sequences of the invention  
 XX are used to evaluate compounds, especially activators or inhibitors of  
 XX bacterial life cycle, for the ability to bind an H. pylori nucleic acid  
 XX sequence. The nucleic acid sequences, and corresponding proteins, are  
 XX also useful for generating vaccines for immunising subjects against H.  
 XX pylori or for use in detecting the presence of Helicobacter species in

a sample. Antisense nucleic acid sequences of these sequences are  
 used to inhibit expression of a gene from Helicobacter species. H.  
 pylori whole genomic DNA was isolated and nebulised to a median size of  
 2000 bp. Purified DNA fragments were blunt-ended and ligated to unique  
 BstXI-linker adapters in 100-1000 fold molar excess. These linkers are  
 complementary to the BstXI-cut PMPX vectors, while the overhang is not  
 self-complementary. Therefore the linkers will not concatenate nor  
 will the cut vector re-ligate itself easily. The linker-adaptor inserts  
 were ligated to each of the 20 PMPX vectors to construct a series of  
 shotgun subclone libraries. The purified DNA samples were then  
 sequenced.  
 Note: The ORF/protein reference number for this sequence was obtained  
 from the related specification, WO9640893.  
 XX  
 XX Sequence 440 BP; 160 A; 68 C; 92 G; 102 T; 18 other;  
 SQ

Query Match 2.8%; Score 39.2; DB 18; Length 440;  
 Best Local Similarity 51.5%; Pred. No. 1.6;  
 Matches 68; Conservative 8; Mismatches 56; Indels 0; Gaps 0;

Qy 962 cttccaagattcactgcccgcacaccccgtaactcaaatgttgccataatcttgctgctc 1021  
 :: | ||| | |||::: | || |||| ::| | ||| | |||| ||  
 Db 418 YVACAAAGCGCGCGSYVYVYAAATTTCCCTCTCKTKTCTAAATTTGGCGTCTT 359  
 Qy 1022 ttttcattccattctctctctgcgaactatcgctgcagacgctgcttatctctgtt 1081  
 ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |  
 Db 358 TTTCAAGCTCTTTTTTTCCTTTTCCAAATAGCGTTTTTTCATCTCTTACAAATTTCTTGT 299  
 Qy 1082 tcttgagtcta 1093  
 || | ||||  
 Db 298 GCTGCGCTTCTA 287

RESULT 14  
 AAS46477  
 ID AAS46477 standard; DNA; 14112 BP.  
 XX  
 AC AAS46477;  
 XX  
 DT 18-DEC-2001 (first entry)  
 XX  
 DE Tumour suppressor gene derived chemically modified sequence #199.  
 XX  
 KW Human; tumour suppressor gene; oncogene; antitumour; cytostatic;  
 KW cancer; tumour; CpG dinucleotide; single-nucleotide polymorphism; SNP;  
 KW cytosine methylation; ds.  
 XX  
 XX Homo sapiens.  
 XX  
 XX WO200168912-A2.  
 XX  
 XX 20-SEP-2001.  
 XX  
 XX 15-MAR-2001; 2001WO-EP02955.  
 XX  
 XX 15-MAR-2000; 2000DE-1013847.  
 XX 06-APR-2000; 2000DE-1019058.  
 XX 07-APR-2000; 2000DE-1019173.  
 XX 30-JUN-2000; 2000DE-1032529.  
 XX 01-SEP-2000; 2000DE-1043826.  
 XX  
 XX (EPIG-) EPIGENOMICS AG.  
 XX  
 XX Olek A, Piepenbrock C, Berlin K;  
 XX  
 XX WPI; 2001-602752/68.  
 XX  
 XX Fragments of chemically modified genes associated with tumour suppressor  
 XX genes and oncogenes, useful in designing primers and probes for  
 XX analysing diseases associated with cytosine methylation state e.g.  
 XX cancer  
 XX

02-JUL-2001; 2001WO-EP07537.  
30-JUN-2000; 2000DE-1032529.  
01-SEP-2000; 2000DE-1043826.  
(EPIG-) EPIGENOMICS AG.  
Olek A, Piepenbrock C, Berlin K;  
WPI; 2002-130909/17.  
Nucleic acid comprising fragment of chemically modified gene, useful  
for diagnosis and treatment of diseases associated with abnormal  
cytosine methylation -  
Claim 1; SEQ ID NO 1415; 32pp + Sequence Listing; German.  
The present invention provides a number of human immune system associated  
genes which are modified by the methylation of cytosines. The sequences  
can be used in the diagnosis and treatment of immune system disorders,  
including eye diseases such as retinopathy, neovascular glaucoma and  
macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
diseases. The present sequence is a gene of the invention.

Sequence 14112 BP; 2722 A; 568 C; 4223 G; 6599 T; 0 other;

```

Sequence 17122 257 4722 67 4222 67 9553 17 0 GENE1
Very Match      2.8%; Score 39.2; DB 24; Length 14112;
Local Similarity 50.0%; Pred. No. 6.8;
tches 98; Conservative 0; Mismatches 98; Indels 0; Gaps 0
1192 tactcaagattgcccctcagctttccgttttgacactcttatggttgcccttcaccgcgt 1251
    || ||| ||| || | ||| || | ||| || | ||| || | ||| || |
3242 tatggaaggcttttagcgctgtttttttttaaggctgttttgttgagcgctttatatgg 3301
1252 ggocctgggccgcgtgtttattgagcacctcgttgacaagcccaagtatactataaatatt 1311
    || | || | || | || | || | || | || | || | || | || |
3302 aggattttagggagtttttagattatttaggagtagggtaaaaatftttataattgt 3361
1312 atagcattagatatacttgtagtctgttttttttatttaagctggtgaataaaataatta 1371
    || | || | || | || | || | || | || | || | || | || |
3362 attgaggtataaaaactgtatataatttctgtgcatgatatatttataaagtcca 3421
1372 ttaaaaaaaaaaaaaa 1387
    || | || | || | || |
3422 ttattataataaaaaa 3437

```

```
DD 342Z UUUUUUUUUUUU 3437  
Search completed: July 23, 2002, 19:06:47  
Job time: 8645 sec
```

26-MAR-2002 (first entry)  
Human immune system associated gene SEQ ID NO: 1415



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 23, 2002, 22:11:17 ; Search time 89.48 seconds  
(without alignments)  
3810.229 Million cell updates/sec

Title: US-08-816-011f-36  
Perfect score: 1388  
Sequence: 1 atggttaataatcaacgctc.....ttattataaaaaaaaaa 1388

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 15 summaries

Database : Issued Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCrUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	68	4.9	7218	1	US-08-232-463-14
2	38.6	2.8	1582	3	US-08-545-196B-10
3	38.6	2.8	1582	3	US-08-545-196B-12
4	36	2.6	1164	2	US-08-794-796-1
5	36	2.6	2621	2	US-08-553-619B-8
6	36	2.6	5852	1	US-07-867-106-2
7	35.8	2.6	1400	1	US-08-464-164-1
8	35.8	2.6	1400	1	US-08-338-057-1
9	35.8	2.6	1400	2	US-08-668-416-1
10	35.4	2.6	581	3	US-08-557-309B-22
11	35.4	2.6	581	3	US-08-834-306-22
12	35.4	2.6	581	3	US-08-993-674A-22
13	35.4	2.6	1303	2	US-08-793-410-11
14	35.4	2.6	3489	2	US-08-728-323A-1
15	35.4	2.6	32207	2	US-08-770-379-20
16	35.4	2.6	32207	4	US-08-757-669A-20
17	35.4	2.6	32207	4	US-09-230-371A-20
18	35.2	2.5	972	3	US-09-286-690-1
19	35	2.5	51259	3	US-08-781-891-209
20	34.8	2.5	289	4	US-09-007-005-17
21	34.8	2.5	289	4	US-09-244-796-17
22	34.6	2.5	991	4	US-09-377-648-7
23	34.6	2.5	1262	4	US-09-227-357-27
24	34.6	2.5	2238	1	US-08-742-011-1
25	34.6	2.5	3645	2	US-08-663-112-1
26	34.2	2.5	852	4	US-09-461-697-1
27	34.2	2.5	1991	2	US-08-415-593-40

28	34.2	2.5	2128	2	US-08-415-593-39	Sequence 39, Appl
29	33.8	2.4	519	1	US-08-339-582-1	Sequence 1, Appl
30	33.6	2.4	732	4	US-08-916-576B-5	Sequence 5, Appl
31	33.4	2.4	756	2	US-08-530-165-1	Sequence 1, Appl
32	33.2	2.4	174	2	US-08-378-235B-8	Sequence 8, Appl
33	33.2	2.4	248	4	US-09-007-005-32	Sequence 32, Appl
34	33.2	2.4	248	4	US-09-244-796-32	Sequence 32, Appl
35	33.2	2.4	277	4	US-09-007-005-3	Sequence 3, Appl
36	33.2	2.4	277	4	US-09-244-796-3	Sequence 3, Appl
37	33.2	2.4	1364	1	US-08-265-087-3	Sequence 3, Appl
38	33.2	2.4	1364	1	US-08-621-493-3	Sequence 3, Appl
39	33.2	2.4	1364	2	US-08-965-688-3	Sequence 3, Appl
40	33.2	2.4	1364	4	US-09-260-173-3	Sequence 3, Appl
41	33.2	2.4	2265	3	US-09-369-618-3	Sequence 3, Appl
42	33.2	2.4	2265	3	US-09-369-617-3	Sequence 3, Appl
43	33.2	2.4	6124	4	US-08-213-419B-3	Sequence 3, Appl
44	33	2.4	306	4	US-09-122-400B-9	Sequence 9, Appl
45	33	2.4	1939	1	US-07-715-751B-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1  
US-08-232-463-14  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: pTgtgt-Fls  
US-08-232-463-14

```

US-08-545-196B-10

Query Match          2.8%; Score 38.6; DB 3; Length 1582;
Best Local Similarity 63.4%; Pred. No. 0.18;
Matches 59; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

Qy 1296 tatottataaatatttatagcatagtgtaactgttatactatgtttttttaaagtgcg 1355
      ||||| | | |||| | | |||| | | |||| | | |||| | | |||| | |
Db 1441 TATCTTCTATATGTTAAAAAGTATATAAATAAAATATTATTTTTTTTTTAAAAAAA 1500

Qy 1356 tgaataaataattattaataaaaaaa 1388
      || |||| | | | ||||| ||||| |||||
Db 1501 AAAAAAAAAAAAAAAAAAAAAAAAAA 1533

RESULT 3
US-08-545-196B-12
: Sequence 12, Application US/08545196B
: Patent No. 6080577
: GENERAL INFORMATION:
: APPLICANT: MELKI, JUDITH
: APPLICANT: MUNNICH, ARNOLD
: TITLE OF INVENTION: SURVIVAL MOTOR NEURON (SMN) GENE: A GENE
: TITLE OF INVENTION: FOR SPINAL MUSCULAR ATROPHY
: NUMBER OF SEQUENCES: 65
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH, LLP
: STREET: PO BOX 747
: CITY: FALLS CHURCH
: STATE: VA
```

```

; ZIF: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/545,196B
; FILING DATE: 19-OCT-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: FARACI, C. J.
; REGISTRATION NUMBER: 32,350
; REFERENCE/DOCKET NUMBER: 2121-110P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8030
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1582 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; US-08-545-196B-12

Query Match 2.8%; Score 38.6; DB 3; Length 1582;
Best Local Similarity 63.4%; Pred. No. 0.18;
Matches 59; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 1296 tatctataaatttatgacattagagatatactgtttattatgttttttataagctg 1355
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1441 TATCTCTTATGTCTTTAAAAAGTATAATAAAAAATATTTAAATTTTTTTTTTAAAAA 1500
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1356 tggataataataattattataaaaaa 1388
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1501 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1533
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
; RESULT 4

```







```

; MOLECULE TYPE: CDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; ORIGINAL SOURCE:
; ORGANISM: Eimeria maxima
; STRAIN: Houghton
; DEVELOPMENTAL STAGE: sporozoite
; IMMEDIATE SOURCE:
; LIBRARY: sporozoite cDNA cloned in Lambda ZAPII
; CLONE: Em70-1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1368
; US-08-338-057-1

```

```

Query Match          2.6%; Score 35.8; DB 1; Length 1400;
Best Local Similarity 69.0%; Pred. No. 1.1; Mismatches 0; Gaps 0;
Matches 49; Conservative 0; Indels 0;

```

```

Qy 1318 tttaggtatactgttatgtttttttaaagctgtggaataaaataattattaaaa 1377
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1326 TCAGCAATGCTTCTTAATATGTTGGTAATCTGCACGACGAATAAATAATAATAAAA 1385

Qy 1378 aaaaaaaaaa 1388
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1386 AAAAAAAAAA 1396

```

```

RESULT 9
US-08-668-416-1
; Sequence 1, Application US/08668416
; Patent No. 5843722
; GENERAL INFORMATION:
; APPLICANT: Tomley, Fiona M.
; APPLICANT: Dunn, Paul P. J.
; APPLICANT: Bumstead, Janene M.
; APPLICANT: Vermeulen, Arno N.
; TITLE OF INVENTION: Coccidiosis poultry vaccine
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Akzo No. 5843722el Patent Department
; STREET: 1300 Piccard Drive, Suite 206
; CITY: Rockville
; STATE: Maryland
; COUNTRY: U.S.A.
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/668,416
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/464,164
; FILING DATE: June 2, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Gormley, Mary E.
; REGISTRATION NUMBER: 34,409
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 258-5200
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1400 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA to mRNA
; HYPOTHETICAL: NO

```

```

; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; ORIGINAL SOURCE:
; ORGANISM: Eimeria maxima
; STRAIN: Houghton
; DEVELOPMENTAL STAGE: sporozoite
; IMMEDIATE SOURCE:
; LIBRARY: sporozoite cDNA cloned in Lambda ZAPII
; CLONE: Em70-1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1368
; US-08-668-416-1

```

```

Query Match          2.6%; Score 35.8; DB 2; Length 1400;
Best Local Similarity 69.0%; Pred. No. 1.1; Mismatches 0; Gaps 0;
Matches 49; Conservative 0; Indels 0;

```

```

Qy 1318 tttaggtatactgttatgtttttttaaagctgtggaataaaataattattaaaa 1377
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1326 TCAGCAATGCTTCTTAATATGTTGGTAATCTGCACGACGAATAAATAATAATAAAA 1385

Qy 1378 aaaaaaaaaa 1388
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1386 AAAAAAAAAA 1396

```

```

RESULT 10
US-08-557-309B-22
; Sequence 22, Application US/08557309B
; Patent No. 5916572
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,309B
; FILING DATE: 14-NOV-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.422
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 581 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-557-309B-22

Query Match          2.6%; Score 35.4; DB 2; Length 581;
Best Local Similarity 61.3%; Pred. No. 0.89;

```

Matches 57; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Qy 1296 tatctataaaattatagcattagctatctgtttatctgttttttttttttaagctg 1355

Db 486 TTTTATATGCTCTTTTATTTGAGAGATGCTCTGTTGTTGTTGTTTTCATCA 545

Qy 1356 tggataaaataattattaaaaa 1388

Db 546 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 578

## RESULT 11

US-08-834-306-22

; Sequence 22, Application US/08834306

; Patent No. 6054135

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Lodes, Michael J.

; APPLICANT: Houghton, Raymond L.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T

; NUMBER OF SEQUENCES: 65

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/834.306

; FILING DATE: 15-APR-1997

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.422C1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 22:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 581 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-834-306-22

## Query Match

Best Local Similarity 2.6%; Score 35.4; DB 3; Length 581;

; Sequence 22, Application US/08834306

; Patent No. 6054135

Matches 57; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Qy 1296 tatctataaaattatagcattagctatctgtttatctgttttttttttttaagctg 1355

Db 486 TTTTATATGCTCTTTTATTTGAGAGATGCTCTGTTGTTGTTTTCATCA 545

Qy 1356 tggataaaataattattaaaaa 1388

Db 546 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 578

## RESULT 12

US-08-993-674A-22

; Sequence 22, Application US/08993674A

; Patent No. 6228372

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Lodes, Michael J.

; APPLICANT: Houghton, Raymond L.

; APPLICANT: Smith, John M.

; APPLICANT: McNeill, Patricia D.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF

; NUMBER OF SEQUENCES: 81

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/993.674A

; FILING DATE: 18-DEC-1997

; CLASSIFICATION: 424

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.422C2

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 22:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 581 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-993-674A-22

## Query Match

Best Local Similarity 2.6%; Score 35.4; DB 4; Length 581;

; Sequence 11, Application US/08793410

Matches 57; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Qy 1296 tatctataaaattatagcattagctatctgtttatctgttttttttttttaagctg 1355

Db 486 TTTTATATGCTCTTTTATTTGAGAGATGCTCTGTTGTTGTTTTCATCA 545

Qy 1356 tggataaaataattattaaaaa 1388

Db 546 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 578

## RESULT 13

US-08-793-410-11/c

; Sequence 11, Application US/08793410

; Patent No. 5955650

; GENERAL INFORMATION:

; APPLICANT: HITZ, WILLIAM DEAN

; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF CANOLA

; TITLE OF INVENTION: AND SOYBEAN PALMITOYL-ACP THIO-

; TITLE OF INVENTION: ESTERASE GENES AND THEIR USE IN

; TITLE OF INVENTION: THE REGULATION OF FATTY ACID

; TITLE OF INVENTION: CONTENT OF THE OILS OF SOYBEAN

; NUMBER OF SEQUENCES: 32

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY

; STREET: 1007 MARKET STREET

; CITY: WILMINGTON

; STATE: DELAWARE

; COUNTRY: USA

; ZIP: 19898

; COMPUTER READABLE FORM:



